Thermography-Assisted Bearing Condition Monitoring

Wael Moussa

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

in Mechanical Engineering

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering

University of Ottawa

Ottawa, Canada

April 2014

Abstract

Despite the large amount of research work in condition based maintenance and condition monitoring methods, there is still a need for more reliable and accurate methods.

The clear evidence of that need is the continued dependence on time based maintenance, especially for critical applications such as turbomachinery and airplane engines. The lack of accurate condition monitoring systems could lead to not only the unexpected failures as well as the resulting hazards and repair costs, but also a huge waste of material and time because of unnecessary replacement due to false alarms and unnecessary repair and maintenance.

Temperature change is a phenomenon that accompanies every dynamic activity in the universe. However, it has not been adequately exploited for mechanical system condition monitoring. The reason is the slow response of current temperature monitoring systems compared to other condition monitoring methods such as vibration analysis. Many references inferred that the change in temperature is not sensible until approaching the end of the monitored component life and even the whole system life (Kurfess, et al., 2006; Randall, 2011; Patrick, et al., March 7-14, 2009).

On the other hand, the most commonly used condition monitoring method, i.e., vibration analysis, is not free from pitfalls. Although vibration analysis has shown success in detecting some bearing faults, for other faults like lubrication problems and gradual wear it is much less effective. Also, it does not give a reliable indication of fault severity for many types of bearing faults.

The advancement of thermography as a temperature monitoring tool encourages the

reconsideration of temperature monitoring for mechanical system fault detection. In addition to the improved accuracy and responsiveness, it has the advantage of non-contact monitoring which eliminates the need for complex sensor mounting and wiring especially for rotating components.

Therefore, in current studies the thermography-based monitoring method is often used either as a distinct method or as a complementary tool to vibration analysis in an integrated condition monitoring system. The main objectives of this study are hence to:

1. Define heat sources in the rolling element bearings and overview two of the most famous bearing temperature calculation methods~~.~~
2. Setup a bearing test rig that is equipped with both vibration and temperature monitoring systems.
3. Develop a temperature calculation analytical model for rolling element bearing that include both friction calculation and heat transfer models. The friction calculated by the model will be compared to that calculated using the pre-defined empirical methods. The heat transfer model is used for bearing temperature calculation that will be compared to the experimental measurement using different temperature monitoring devices.
4. Propose a new in-band signal enhancement technique, based on the synchronous averaging technique, Autonomous Time Synchronous Averaging (ATSA) that does not need an angular position measuring device. The proposed method, in addition to the Spectral Kurtosis based band selection, will be used to enhance the bearing envelope analysis.
5. Propose a new method for classification of the bearing faults based on the fault severity and the strength of impulsiveness in vibration signals. It will be used for planning different types of tests using both temperature and vibration methods.
6. Develop and experimentally test a new technique to stimulate the bearing temperature transient condition. The technique is supported by the results of finite element modeling and is used for bearing temperature condition monitoring when the bearing is already running at thermal equilibrium condition.

Acknowledgments

Firstly I would like to express the great admittance to God for giving me the health and the power to finish this work.

I also want to express great appreciation to my sponsor, Egyptian Armed Forces, for their continuing support despite the challenges the country is passing through.

I also want to express my great gratitude to my supervisor Prof. Ming Liang for his support and guidance and the great thankfulness to the members of PhD thesis committee.

I also want to express my thankfulness to my colleagues in the research group and to the machine workshop staff in the University of Ottawa, Mr. John Perrins, Mr. Léo Denner and Mr. Stan who helped me during the experimental work of this study.

Finally, I cannot find the words to express my gratitude to my family members who are always the real backup and the continuous motive in my life.

Table of Contents

ABSTRACT .... **I**

ACKNOWLEDGMENTS .... **III**

TABLE OF CONTENTS .... **IV**

LIST OF FIGURES .... **IX**

LIST OF TABLES .... **XV**

NOMENCLATURE .... **XVII**

LIST OF ACRONYMS .... **XXII**

1. INTRODUCTION **1**

1.1 CHALLENGES IN CONDITION MONITORING 1

1.2 CONDITION MONITORING METHODS 3

1.3 TEMPERATURE-BASED MONITORING METHODS 4

1.4 VIBRATION-BASED MONITORING METHODS 5

1.5 THESIS OUTLINE 5

2. LITERATURE REVIEW **10**

2.1 HISTORICAL PERSPECTIVE 10

2.2 FAULT DETECTION AND DIAGNOSIS METHODS FOR ROLLING ELEMENT BEARINGS 11

2.2.1 Vibration analysis methods 12

2.2.1.1 Envelope window selection methods 13

2.2.1.2 Filtering methods 14

2.2.2 Temperature monitoring methods *15*

2.3 FRICTION IN ROLLING CONTACTS 16

2.3.1 Energy losses due to friction *17*

2.3.2 Lubricant film thickness in Elasto-Hydrodynamic Lubrication (EHL) *18*

2.3.3 Friction reduction caused by vibration *19*

2.4 SUMMARY 20

3. BEARING SURFACE TEMPERATURE 22

3.1 HEAT SOURCES IN ROLLING ELEMENT BEARINGS 23

3.2 EMPIRICAL FORMULAS FOR BEARING HEAT GENERATION RATE 25

3.2.1 Palmgren model *26*

3.2.2 SKF model *28*

3.3 ANALYTICAL SOLUTION FOR BEARING HEAT GENERATION RATE 29

3.3.1 Speeds and forces for rolling bearing contacting elements *30*

3.3.1.1 Kinematic solution for bearing speeds 30

3.3.1.2 Complete dynamic solution 32

3.3.2 Surface friction shear stress in rolling element/raceway contact *41*

3.3.2.1 Boundary lubrication 42

3.3.2.2. Hydrodynamic lubrication 43

3.3.2.3 Partial or mixed lubrication regime 46

3.3.2.4 Lubricant film thickness and ratio for solid-solid contact 48

3.3.3 Viscous drag force *52*

3.3.4 Total heat generation rate *53*

3.4 ROLLING ELEMENT BEARING TEMPERATURES AND HEAT TRANSFER MODELING 56

4.CALCULATION OF BEARING TEMPERATURE 58

4.1 BEARING TEMPERATURE MODEL 58

4.1.1 Bearing dry friction calculation -- Module 1 59

4.1.2 Bearing hydrodynamic friction calculation – Module 2 63

4.1.3 Bearing temperature calculation – Module 3 65

4.1.3.1 Theoretical model, lumped thermal capacitance 65

4.1.3.2 Equations of heat transfer modes and thermal resistances............................. 69

4.1.3.3 Computational algorithm 71

4.2 HEAT GENERATION MODEL VERIFICATION.. 74

4.2.1 Verification by the Palmgren model.. 75

4.2.2 Verification by the SKF model 76

4.2.2.1 Rolling friction moment M_r 77

4.2.2.2 Sliding friction 77

4.2.2.3 Seal friction.. 79

4.2.3 Comparison of the results obtained by the Palmgren, SKF and the proposed friction models.. 80

4.2.3.1 Comparison at the thermal equilibrium condition 80

4.2.3.2 Comparison at the thermal transient condition 83

4.3 BEARING TEMPERATURES, HEAT TRANSFER MODEL VERIFICATION 84

4.3.1 Test equipment.. 85

4.3.2 Surface temperature measurement devices 86

4.3.3 Normal bearing tests 89

5. VIBRATION-BASED BEARING FAULTDETECTION:ENVELOPE ANALYSIS ... 93

5. 1 ENVELOPE WINDOW SELECTION 94

5.1.1 Application of EW selection based on STFT-SK method *97*

5.2 IN-BAND SIGNAL ENHANCEMENT... 102

5.2.1 Deterministic signal removal using a linear prediction filter *104*

5.2.2 Fault signal enhanced by autonomous time synchronous averaging *104*

5.3 VALIDATION OF THE PROPOSED ALGORITHM FOR BEARING FAULT DETECTION..................... 107

5.3.1 Validation using simulated signal *107*

5.3.1.1 Slippage effect removal. 108

5.3.1.2 Discrete noise removal. 111

5.3.2 Validation using experimental data *115*

5.4 SUMMARY OF THE BEARING VIBRATION ANALYSIS METHOD 119

6. BEARING FAULT DETECTION 121

6.1 BEARING FAULTS CLASSIFICATION 122

6.2 FAULT SEEDING 124

6.2.1 Impulsive faults: inner and outer race spalls *124*

6.2.2 Non-impulsive faults: lubricant deficiency *126*

6.3 TEMPERATURE-BASED BEARING CONDITION MONITORING 127

6.3.1 Effects of impulsive faults on bearing transient temperature *128*

6.3.1.1 Impact force, F_{imp} 129

6.3.1.2 Fault impact force distribution 131

6.3.2 Effects of non-impulsive fault on temperature *135*

6.4 EXPERIMENTAL RESULTS 137

6.4.1 Tests of bearings with impulsive faults *138*

6.4.1.1 Short time test 139

6.4.1.2 Extended tests 143

6.4.1.3 Matured fault test 145

6.4.2 Tests of bearings with a non-impulsive fault.. 148

6.4.3 Summary of experimental results 152

7. STIMULATION OF BEARING THERMAL TRANSIENT BEHAVIOR.................... 153

7.1 TRANSIENT STIMULATION SYSTEM MODELING... 154

7.2 EXPERIMENTAL RESULTS 159

8. SUMMARY, CONCLUSION AND FUTURE WORK.. 161

8.1 SUMMARY 161

8.2 CONCLUSIONS 162

8.3 FUTURE WORK.. 165

. REFERENCES..166

APPENDIX.. 180

LUBRICANT RHEOLOGY 180

A.1 TEMPERATURE EFFECT ON LUBRICANT PROPERTIES 181

A.2 PRESSURE VISCOSITY RELATIONSHIP 183

A.3 SHEAR RATE EFFECT.. 185

List of Figures

Figure 1.1. Detection horizons of different condition monitoring methods (Patrick, et al., March 7-14, 2009) 3

Figure 3.1 Radial ball bearing internal dimensions and speeds. 31

Figure 3.2 General motion descriptions for rolling element in a radial ball bearing. 33

Figure 3.3 Radial ball bearing contact areas and radii of curvature. 36

Figure 3.4 Curvature radii for two curved contacting hemisphere bodies. 37

Figure 3.5 Distribution of: (a) sliding velocities, and (b) friction shear stress, τ, at any point, (*x*, *y)* of the elliptical area of contact (Harris, et al., 2007) 43

Figure 3.6 Lubricant tangential speed distributions for a) fixed raceway, and b) rotating raceways. 45

Figure 3.7. a) The mixed lubrication between two rough surfaces, and b) details of a single asperity showing different planes: plane (1) is the mean plane for the rough surface, plane (2) is the mean plane for the summits, plane (3) is the largest height before possible solid to solid contact and plane (4) is the summits height plane 52

Figure 3.8 Calculation scheme for the total friction for lubricated rolling element bearing. 55

Figure 4.1 Normal stress distribution in a bearing rolling contact area 62

Figure 4.2 The heat transfer network for the radial ball bearing, its housing and the shaft 67

Figure 4.3 Flowcharts for bearing temperature calculation 73

Figure 4.4 Heat generation rates obtained by the three different models: the Palmgren model, the SKF model and the proposed model, calculated at low loading condition, 10 to 50 N radial loads. 82

Figure 4.6 Effect of speed on different heat generation sources as defined by the proposed model 83

Figure 4.7 Effect of the load on the different heat generation sources as defined by the proposed model. 83

Figure 4.8 Bearing temperature test rig with different loading setups: (a) single load disc, (b) two load discs, and (c) three load discs. 88

Figure 4.9 1-AC motor (up to 20,000 rpm), 2-Two-step 1566 steel shaft, 3-Anti-vibration flexible couplings (2), 4-Radial load (Disc), 5-Shaft-Hub Locking Device (SHLD).6-Bearing and bearing Housings (2 each), 7-Motor base, 8-Sliding base, 9-HD-810 dynamometer and 10-Sliders (4), 11-FLIR E40 thermal camera (Not shown), 12-Infrared sensor (Not shown) and 13- thermocouples at bearing outer race (2) (Not shown) 88

Figure 4.10 Comparison between the temperature rise curves for inner and outer races surfaces as measured using different measuring devices and as calculated using the heat transfer model in accordance with the heat generation rate calculated by SKF model, Palmgren model and the proposed model 90

Figure 4.11 Comparison between the temperature rise curves for inner and outer races surfaces as measured using different measuring devices and as calculated using the adaptive bearing temperature model. 92

Figure 5.1 Window length selection algorithm for STFT-based SK method. 98

Figure 5.2 Raw signal acquired by an accelerometer for a bearing with an outer race fault running at steady speed of 6000 rpm 99

Figure 5.3 The time-frequency representation of the bearing signal obtained using the STFT method 100

Figure 5.4 (a) The SK values as calculated at any frequency and using different window lengths, and (b) a zoomed section at maximum SK 101

Figure 5.5 Vibration-based bearing fault detection scheme. 109

Figure 5.6 Simulated bearing fault at different shaft speeds: a) at 8 Hz, and b) at 100 Hz. 110

Figure 5.7 Simulated bearing fault with the slippage effect. 110

Figure 5.8 Slippage effect on bearing envelope spectrum appeared in the envelope analysis results (a-f), enhanced envelope analysis results by the TSA method (g-I) and enhanced envelope analysis by the ATSA method (j-l). Knowing that slippage ratio is zero for a, d, g and j, 2% for b, e, h and k and 5% for c, f, i and l, the first raw represents the effect of the slippage on a clean signal, i.e. zero noise while for all other simulated signals a white Gaussian noise of 1 dB is added. 112

Figure 5.9 Envelope spectrums for the simulated interfered bearing signal with a simulated gear meshing signal when considering a ± 2% bearing slippage (a) and ± 5% bearing slippage (b). (c) and (d) represent the envelope spectrums by the basic envelope analysis method, (e) and (f) represent the reconstructed signals using the TSA method while (g) and (h) are the envelope spectrums for the reconstructed signals (e) and (f), respectively. Plots (i) and (j) represent the reconstructed signal using ATSA method while (k) and (l) are the envelope spectrums for the reconstructed signals (i) and (j), respectively... 116

Figure 5.10 Inner race faulty bearing raw vibration signal while shaft speed is 100 Hz........... 117

Figure 5.11 Bearing signal alignments using the TO peak localization: (a) the windows of the raw signal before alignment, and (b) the aligned windows. 118

Figure 5.12 The resulting envelope spectrum for an inner race faulty bearing when using (a) the original raw signal, and (b) the signal enhanced by the ATSA method. 119

Figure 6.1 a) The EDM electrode, b) 0.6 mm electrode tip, and c) fixtures for both bearing inner and outer races 125

Figure 6.2 a) Three faulty inner races with spalls of 0.6, 1 and 1.4 mm wide respectively, b) three faulty outer races with spalls of 0.6, 1 and 1.4 mm wide respectively, and c) close-up view of the 1-mm-wide outer race spall 125

Figure 6.3 Bearing spall fault shape. 125

Figure 6.4 Lubricant quantity measuring tools 126

Figure 6.5 Ball impact to the fault edge (scale 10:1 of real dimensions) 129

Figure 6.6 The forced non conservative oscillating system representing the motion of the rolling element between the bearing races with the fault effect. 132

Figure 6.7 Finite element modeling of the bearing friction under decaying oscillating load using ANSYS software. 134

Figure 6.8 Modeling of the contact pair for both inner and outer rings with the rolling element using ANSYS software 135

Figure 6.9 Friction shear stresses at the contact area calculated using the ANSYS FE software 135

Bearing condition 141

Figure 6.10 Snapshots taken by a FLIR-E40 thermal camera for the side surface of the inner race of three bearings at three different conditions, healthy, faulty outer race and faulty inner race, recorded 5, 10, 50 and 100 seconds into the tests running at 100 Hz shaft speed. 141

Figure 6.11 The temperature rise curves measured for 300 seconds for the healthy bearing, two bearings with different sized spalls, and the calculated curve for the healthy bearing. 142

Figure 6.12 The temperature rise curves measured for 300 seconds for the healthy bearing, two faulty bearings (one with an inner race spall and the other with an outer race spall of the same size, 1 mm in depth), and the calculated curve for the healthy bearing. 142

Figure 6.13 Temperature rise differencesbetween: (a) the temperatures of the reference healthy bearing and the bearing with an inner race fault, and (b) thetemperatures of the reference healthy bearing and the bearing with an outer race fault. 143

Figure 6.14 Extended test results of healthy and faulty inner race bearings. 144

Figure 6.15 bearing outer race fault at different maturity levels: (a) after 10^5 shaft cycles, (b) after 10^6 shaft cycles 146

Figure 6.16 Envelope spectra at fault characteristic frequency after (a) 100000 shaft cycles, and (b) 6,400,0000 shaft cycles 147

Figure 6.17 Temperature rise curves for a healthy bearing as a reference and a bearing with outer race faults with different maturity levels (a) after 100000 shaft cycles, and (b) after 6,400,0000 shaft cycles. 148

Figure 6.18 Temperature rise curves for different lubricant quantities in the same bearing and running at the same operating conditions. 150

Figure 6.19 Vibration spectrums using FFT method for the bearing at six different lubricating conditions, different lubricant quantities.. 151

Figure 6.20 Vibration envelope spectrums using the ATSA method in 5.2 for the bearing at six different lubricating conditions, different lubricant quantities. 151

Figure 7.1 The thermal network for the TSS. 155

Figure 7.2 Bearing test rig with a surface heater installed.. 158

Figure 7.3 Temperature rise curves calculated at points No. 7 and point No. 4 from Figure 6.1, using the lumped mass method model 158

Figure 7.4 Finite element analysis results using ANSYS at the end of the energizing time that shows the heat flow direction is in negative z direction 159

Figure 7.5 Experimental results for both the initial transient and stimulated transient conditions for both healthy and faulty bearings.. 160

Figure A.1 Lubricant viscosity versus operating temperature using Vogels equation. 182

Figure A.2 Lubricant specific heat in (KJ/Kg.K) versus temperature. 182

Figure A.3 Lubricant thermal conductivity (W/mK) versus operating temperature 183

List of Tables

Table 4.2 Parameters of the test bearing (Type:, SKF 6202-16). 74

Table 4.3 Viscous friction torque for different speeds 75

Table 4.4 Friction torque for different loading cases obtained by the Palmgren Model 76

Table 4.5 Heat generation rates obtained by the Palmgren model (in watts) 76

Table 4.6 Rolling friction moments (in N.mm) obtained using the SKF model for different loading and speed conditions. 77

Table 4.7 Sliding moments (in dyne.mm) obtained using the SKF model for different loading and speed conditions. 79

Table 4.8 Heat generation rates (in watts) obtained using the SKF model for different loading and speed conditions. 79

Table 4.9 Constants for calculating seal friction moment for single row deep groove ball bearing (Type: SKF 6202-2RSH) 79

Table 4.10 Main features of the E-40 thermal camera and OS211 infrared sensor 87

Table 5.1 Selected EW bandpass filter parameters: center frequencies and upper and lower cut off frequencies for different bearing conditions based on the SK method (shaft frequency = 100 Hz) 102

Table 5.2 Envelope spectrum enhancement, increase of SNR, for different bearing slippage ratios as depicted in Figure 5.8 (d-l). 111

Table 5.3 Enhancement of SIR for the envelope spectrums expected using three different methods and at two different ratios of bearing slippage 115

Table 6.1 Dimensions of the seeded bearing spalls. 126

Table 6.2 Areas of contact for steady state loading and with impact oscillating force added calculated by the ANSYS finite element software bearing friction using ANSYS software. 134

Table 6.3 Different lubricant levels used in the temperature and vibration tests 149

Nomenclature

A	Contact area for rolling element and bearing ring
A_i	Contact area for rolling element and inner ring
A_o	Contact area for rolling element and outer ring
a	Semi-major axis for elliptical area of contact
a^*	Contact area dimensionless semi-major axis
AC	Area of contact
b	Semi-minor axis for elliptical area of contact
b^*	Contact area dimensionless semi-minor axis
b_w	Bearing width
Bi	Biot Number
c	Load exponent in Palmgren model
c_v	Velocity direction coefficient for friction shear stress calculation
c_d	Lubricant viscous drag coefficient
$\mathbb{C}$	Damping coefficient of material
$\boldsymbol{c}$	Specific heat coefficient
C_s	Bearing static load rating
D	Rolling element diameter
D_b	Ball diameter
d_m	Bearing mean or pitch diameter
D_i	Bearing inner race diameter
D_o	Bearing outer race diameter
d_s	Bearing shoulder diameter
d_{sh}	Shaft diameter
dm	Unit of mass
D_{sum}	Summits density in a surface unit area
E	Elasticity stiffness coefficient
$\mathbb{E}$	Probability expectation operator

$\bar{E}$	Equivalent elasticity
F	Bearing applied load
F_{im}	Impulse force
F_f	Friction force
F_v	Viscous drag force
f_0	Bearing viscous friction coefficient, Palmgren model
f_1	Bearing load friction coefficient, Palmgren model
F_r	Bearing radial load
F_s	Sampling frequency
f_r	Shaft rotational frequency
f	Frequency, Hz
g	Gravitational acceleration
G	Heat generation source symbol
$\mathscr{G}$	Material dimensionless parameter
G_{rr}	SKF bearing rolling friction
G_{sl}	SKF bearing sliding friction
H	Heat generation or dissipation rates in watts
h	Lubricant film thickness
h_v	Coefficient of heat convection
h_{cen}	Lubricant film plateau thickness
H_{RBC}	Bearing raceway contact generated heat
H_{drag}	Lubricant viscous drag generated heat
H_0	Dimensionless minimum lubricant film thickness
H_{cond}	Heat transferred by conduction
H_{conv}	Heat transferred by convection
H_{rad}	Heat transferred by radiation
$\mathscr{H}$	Total momentum, angular and linear.
i	Subscript for bearing inner race
K	Kurtosis
$\mathscr{k}$	Thermal conductivity

L	Bearing load
l	Flow characteristic length in Reynolds No. definition
ℓ	Length between two points with conduction heat transfer
m	Mass unit
M	Total bearing friction moment
M_l	Friction moment due to applied load, Palmgren model
M_v	Viscous friction moment, Palmgren model
M_{hys}	Hysteresis friction moment
M_{de}	Friction torque due to deviation from ideal design
M_{rr}	Rolling friction moment, SKF model
M_{sl}	Sliding friction moment, SKF model
M_T	Friction moment resulted from temperature change effect
M_c	Friction moment due to cage sliding
M_{gyr}	Friction moment due to gyroscopic effect
M_s	Friction moment due to sliding at raceway contact
m	Heat convection coefficient parameter
m	Index for bearing fault simulated signal impacts
n	Rotational speed in revolutions per minute
N	Index for bearing signal windowing
o	Index for bearing outer race
P	Applied pressure
Pr	Prandtl Number
$\mathscr{P}_i$	Internal energy for bearing component i.
Q	Bearing normal load
$\bar{Q}$	Dimensionless loading parameter in lubricant film thickness formula
R	Symbol for radiation
R	Position vector
Re	Reynolds Number
R_s	Summit radius
r	Radius of curvature

s	Specific gravity
S	Standard deviation for bearing rough surface summits height distribution
$\mathcal{S}$	Cross section area through which heat flows, radiation and convection
t	Time in seconds
T	temperature
T_{HE}	Heater energizing time
U	Relative speed at contact area points
$\bar{U}$	Dimensionless speed parameter in lubricant film thickness formulas
v	Velocity
X	Bearing axial direction
Y	Bearing longitudinal direction
Z	Bearing vertical direction
$\bar{Z}_s$	The distance between bearing summits and surface mean planes
α	Bearing angle of contact
α_{bw}	Band width parameter, surface roughness
α_{pv}	Lubricant pressure viscosity coefficient
β	Rolling element pivoting angle
$\beta^{'}$	Rolling element spinning angle
γ	bearing internal dimension ratio
γ_s	Seal friction exponent as defined in SKF earing friction model
δ	Bearing contact area deflection
δ^*	Dimensionless deflection for contact area
δ_0	Deflection at steady state loading
ε	Emissivity coefficient
η	Lubricant dynamic viscosity
η_0	Lubricant dynamic viscosity at ambient pressure and temperature
Λ	Lubricant parameter
μ	Dynamic coefficient of friction
μ_{sl}	SKF bearing sliding friction coefficient
μ_{bl}	SKF bearing lubricant property factor

μ_{EHL}	Friction coefficient for full fluid film
ν_A	Poisson ratio for material A
ν_0	Kinematic viscosity at ambient pressure and temperature
ξ	Lubricant weight ratio to the bearing free space
ρ	Density
σ	Contact area normal stress
σ_s	Summits height standard deviation
τ	Tangential shear stress
τ_f	Friction shear stress
τ_N	Newtonian shear stress
τ_{lim}	Limiting shear stress
ϕ	Bearing mass element local angle
ϕ_{bl}	SKF bearing lubricant weighting factor
ψ	Bearing azimuth angle
ω_i	Bearing inner race angular velocity
ω_o	Bearing outer race angular velocity
ω_m	Bearing mean angular velocity
ω_s	Shaft angular velocity
ω_R	Rolling element rolling angular velocity
$\omega_{x'}$	Rolling element sliding angular velocity in local axial direction
$\omega_{y'}$	Rolling element sliding angular velocity in local pivoting direction
$\omega_{z'}$	Rolling element spinning angular velocity
ω_{si}	Rolling element spinning angular velocity at inner race
ω_{so}	Rolling element spinning angular velocity at outer race
$\sum\rho$	Curvature sum for two contacting bodies

List of Acronyms

ATSA	Autonomous Time Synchronous Averaging
AFB	Anti-friction bearing
BEA	Bearing Envelope Analysis
BEM	Boundary Element Method
BPFI	Ball pass frequency for bearing inner race fault
BPFO	Ball pass frequency for bearing outer race fault
BTTS	Bearing Transient Thermal Signature
CF	Cage frequency
CBM	Condition Based Maintenance
CS	Cyclostationary
EW	Envelope Window
EA	Envelope Analysis
EHD	Elasto-Hydrodynamic
EHL	Elasto-Hydrodynamic Lubrication
FEM	Finite Element Method
FFT	Fast Fourier Transform
GW	Greenwood Williamson Model
HFR	High Frequency resonance
IF	Impulsive Faults
IFC	Impulsive Fault with Consistent Vibration Signature
IFT	Impulsive Fault with Temporary Vibration Signature
LLS	Linear Least Square
LPF	Linear Prediction Filter
LST	Laser Surface Texturing
LDR	Levinson–Durban Recursion algorithm
MG	Multi Grid
MEMS	Micro-Electro-Mechanical Systems
NDT	Non-Destructive Testing

NIF	Non-Impulsive Fault
PDF	Probability Density Function
RBC	Ball-Raceway Contact
SK	Spectral Kurtosis
SNR	Signal to Noise Ratio
STA	Synchronous Time Averaging
STFT	Short Time Fourier Transform
SVF	Scarce-Vibration Faults
TDA	Tear Down Analysis
TDA	Time Domain Averaging
TO	Teager Operator
TSS	Transient Stimulation System
TTS	Transient Thermal Signature
URL	Useful Remaining Life
WT	Wavelet Transform
WVD	Wigner-Ville Distribution
WVT	Wigner-Ville Transform

Chapter 1

Introduction

1.1 Challenges in Condition Monitoring

The increasingly complex mechanical systems calls for more robust and reliable maintenance systems The current maintenance systems can be classified into three main types or strategies (Randall, 2011):

1. Run-to-Break,
2. (Time-Based) Preventive Maintenance, and
3. Condition Based Maintenance.

Although condition based maintenance has many advantages, the first two, especially the time-based strategy, are still in use. There are three main reasons why the condition based maintenance strategy has not replaced the other two in many mechanical and engineering systems.

The first reason is related to industrial acceptance. Many condition monitoring systems are costly in the development of both hardware and software, time-consuming in implementation, and knowledge-demanding in applications. In many cases, such investments and efforts may have to be repeated for each and every vulnerable spot in the system. Hence many companies are very reluctant to adopt the condition based maintenance strategy.

The second reason is the detection certainty. Although condition monitoring systems have been used successfully for a long time, they are still not fully reliable to be used for many critical applications or complex systems. In the case of critical applications such as turbojet engines for civilian and military airplanes, missiles and huge-sized turbo-machinery, a fault could develop very rapidly if it is not detected in a timely manner and can cause a catastrophic breakdown. On the other hand, false alarms due to the artifact or spurious fault signatures can result in a huge waste of time and manpower especially for large and complex systems.

The third reason is the lack of efficient systems that can assess the fault severity and the remaining useful life (RUL). As a result, very conservative maintenance actions are taken leading to a waste of material and time due to unnecessary replacement of system components. For the above reasons, the improvements of a new condition monitoring system should target the following aspects:

1. Simplicity – a practical monitoring system should be inexpensive and fast to develop and implement, easy to use and versatile to be deployed to multiple vulnerable system spots.
2. Certainty – the detection results should be reliable with minimal undetected faults and false alarms.
3. Responsiveness – this can be measured by detection horizon, i.e., the elapsed time between the first detection of a fault and the resultant mechanical failure (Watson, et al., 2007). Obviously, the longer the detection horizon is, the more responsive the system will be.
4. Accuracy in fault severity assessment – this is critical to avoid the excessive maintenance activities.

1.2 Condition Monitoring Methods

A condition monitoring system may require one or more fault detection methods. The main methods used for condition monitoring systems are mentioned below. Multiple methods are often integrated in a system to improve the fault detection certainty and reduce false alarms.

1. Vibration analysis, either traditional methods or more sophisticated multi-spectral analysis.
2. Oil analysis monitoring.
3. Temperature monitoring.
4. High frequency analysis including acoustic emission and stress wave analysis.

Each of the above methods may be suitable for some applications but not for others. The choice of the method depends on the system to be monitored. Also, one of the important factors that differentiates these methods and affects the choice of the suitable method for each system is the detection horizon or the capability of early fault detection. Thus far, it has been believed that temperature monitoring has the shortest detection horizon among all the existing condition monitoring methods (Kurfess, et al., 2006; Randall, 2011; Watson, et al., 2007). Figure 1.1 shows the ranking of different condition monitoring methods in terms of the detection horizon length (Patrick, et al., March 7-14, 2009).

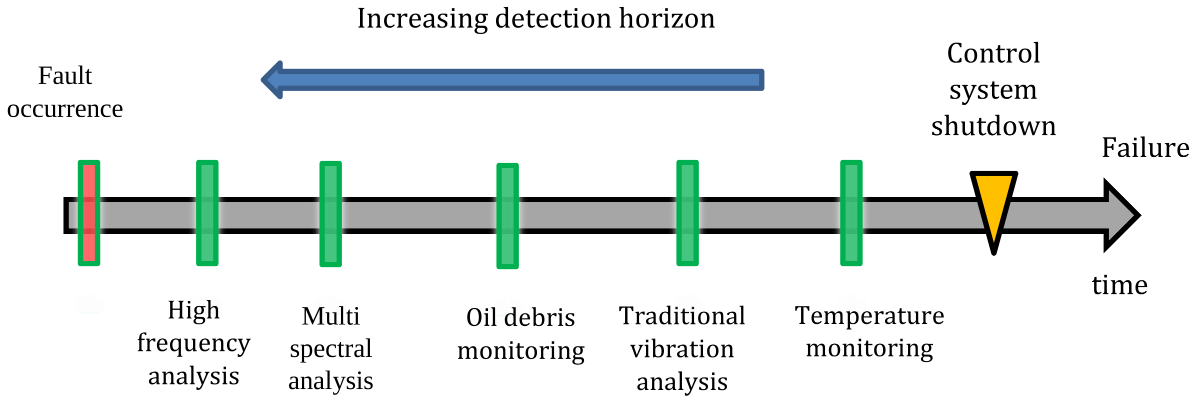


Figure 1.1. Detection horizons of different condition monitoring methods (Patrick, et al., March 7-14, 2009)

1.3 Temperature-Based Monitoring Methods

Heat emission is a phenomenon that accompanies almost every dynamic activity. However, using temperature monitoring for mechanical fault detection has been thought to be less effective than other methods, as shown in the above section. This is mainly because of the misperception that a temperature monitoring system has the shortest detection horizon among all the systems listed in Figure 1.1. However, the detection horizon of temperature based monitoring can be substantially longer than those of its peers depending on:

1. The faults to be detected (e.g., the temperature based method is much more effective than the vibration method in detecting the lack of lubrication and hence much longer detection horizon, though it falls short in the detection of physical bearing damages) and the use of the bearing temperature rise curve as a health assessment method instead of using equilibrium temperature of bearing surface. In this trend the temperature rise curve can be considered as a bearing transient thermal signature, BTTS, that differs according to bearing condition.
2. The understanding of the differences of the heat generation mechanisms due to the normal operation and malfunction of the target system, and
3. The availability of temperature sensing devices with improved accuracy and response time those are able to detect the difference between bearing temperature rise curves (transient thermal signatures) in different bearing conditions.

As rolling element bearing is a main component of many of mechanical systems and its faults are involved in a high percentage of rotary systems failures, a reliable bearing condition monitoring system that can meet the requirements mentioned in section 1.1 is highly desirable. Temperature monitoring can play a crucial role in such a system. To this end, the heat generation mechanism during the operations of the healthy bearings has to be well understood. The effect of bearing fault on heat generation should also be analyzed. The success also relies on accurate measurements and analyses of the temperature rise behavior for both healthy and faulty bearings using a high response temperature sensing device, e.g., a thermal camera and infrared sensor.

1.4 Vibration-Based Monitoring Methods

Vibration-based condition monitoring systems have been used for bearing fault detection over several decades. However, they are effective mainly in the detection of physical bearing damages such as cracks and spalls, and may not be easily used to detect several other undesirable conditions, in particular the lack of lubrication which eventually leads to physical damages of the bearing components. Even though, the vibration-based methods are still the most popular approaches whereas temperature readings are used only for complementary purposes. The most popular bearing assessment method, the envelope analysis, has proved effective for detecting localized bearing faults. Several studies have been carried out to enhance the envelope analysis method in two main directions:

1. Optimization of envelope window selection
2. In-band filtering for noise removal and signal enhancement

The two above directions will be examined in this study for enhanced envelope analysis for bearing fault detection and health assessment.

1.5 Thesis Outline

This thesis focuses mainly on exploiting the bearing heating phenomenon in the condition monitoring field. Instead of the traditional methods that were depending on bearing equilibrium temperature, for bearing health assessment and fault detection, the changes to Bearing Transient Thermal Signature (BTTS) will be monitored by a thermal camera and infrared sensors,. Also, the vibration analysis methods are concerned due to its effectiveness in detection localized damage.

In Chapter 2 a literature review covering both the history and recent studies of both vibration and temperature monitoring methods is presented. The review includes also the studies in the fields of friction in rolling contact, the vibration effect on friction and their applications.

In Chapter 3, the theoretical base behind bearing heat generation mechanism and the methods of bearing temperature calculation are introduced. Both empirical and analytical solutions for the bearing temperature problem are presented in this chapter. .

Chapter 4 reports a mathematical model that calculates bearing temperatures during thermal transient condition and at different bearing health condition. The model calculates the bearing temperature through two main steps: the first step is the calculation of bearing friction and total rate of heat generation due to bearing operation while the second step is the calculation of bearing temperatures using a heat transfer model. In the first step the friction is calculated by the summation of two main parts: dry and hydrodynamic friction. The last part also is combined of two parts: hydrodynamic shear and viscous drag frictions. The model is able to calculate the change in friction due to the change of bearing condition such as the presence of localized fault, lubricant lack and lubricant properties deterioration. The proposed model is verified by comparing its friction results with the friction results calculated at the same operating condition and for the same bearing, using two of the most popular bearing friction empirical models, SKF and Palmgren models. In the second step a heat transfer network is used to calculate the temperature of different bearing components during transient condition. The results of the calculated bearing temperatures resulted from the model second step are compared to the experimental results recorded by temperature monitoring test rig that is setup for this purpose.

In chapter 5 the bearing fault detection using BTTS monitoring is examined and verified through several steps. Firstly, new definitions for bearing faults depending on their vibration signature are proposed, Impulsive and Non-Impulsive Faults. According to the proposed new definitions two sets of faulty bearings with the two types of faults are setup. The faulty bearings are then used for verifying theoretical results for both temperature and vibration-based bearing fault detection methods.

Secondly, a new vibration analysis technique is proposed and tested for both simulated and real data extracted by a bearing test rig with different faulty bearings. The vibration method is used for showing the difference between the bearing faults as defined by the new definition and to compare the effectiveness of the vibration analysis methods with the temperature analysis methods with respect to different types of bearing faults, Impulsive and Non-Impulsive faults, and at different stages of fault life, e.g. incipient and matured stages.

Thirdly, the effect of bearing faults of the two types on the BTTS is studied. The presence of localized fault affects the average area of contact and the resulted dry friction, this effect is studied using finite element modeling while the lack of lubricant will affect the viscous drag friction and is calculated using Harris formula for viscous drag friction.

Fourthly, two groups of experimental test, short and extended time tests are introduced for the bearings with different conditions. The tests include the monitoring of both vibration envelope spectrum using the Enhanced proposed method and temperature monitoring for the BTTS. The results of experimental tests are compared to theoretical results and used to show the effectiveness of each technique in different bearing conditions. The results of this chapter shows the benefit of using both techniques in bearing condition monitoring as each technique can be efficient when the other has difficulties in bearing fault detection.

In order to exploit the rich thermal transient stage in bearing fault detection a new technique for stimulating the thermal transient condition during bearing operation is proposed and implemented in Chapter 6. An external heat source is used to change the bearing thermal condition from equilibrium to transient. The setup of the transient condition stimulating system is assisted by finite element modeling and then verified using experimental results using bearing temperature monitoring test rig. Also the usage of both methods can give higher fault detection certainty and less error probability.

Chapter 7 concludes the thesis and provides suggestions for future work. A flow chart that depicts the whole thesis structure is shown in Figure 1.2.

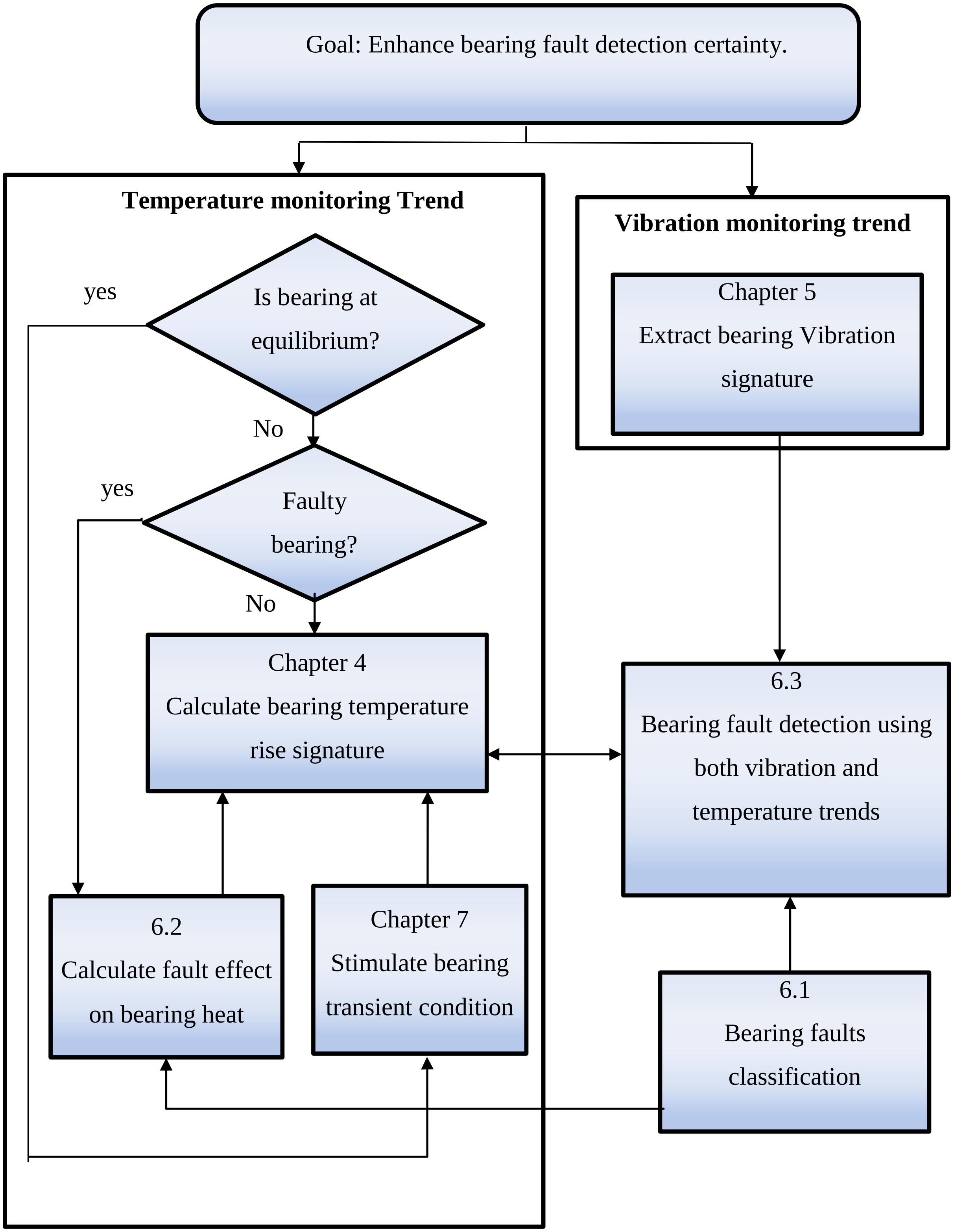


Figure 1.2 Thesis flow chart.

Chapter 2

Literature Review

2.1 Historical Perspective

Rolling elements bearing is one of the successes to understand and control friction phenomenon. From the early discovery of fire initiation to the microscopic friction measuring and control studies, there is a large number of discoveries and applications thanks to the growing understanding of this phenomenon. However, the understanding of this phenomenon is still not complete. One of the first human successes related to friction phenomenon is back in history for thousands of years and referred to The Egyptians, around 2400 BC, who were the first to use a lubricant, water, to reduce the force required to drag carriages used to transfer huge stones and statues. Instead of lubricant, the Assyrians in 1100 BC employed rollers under their carriages to do the same work with less manpower (Harris, et al., 2007).

The discovered fragments that may be related to the oldest rolling element bearing are thought to be these found in Lake Nemi, Italy, in 1928. It was speculated that these fragments were part of a thrust ball bearing that was used in a rotatable statue (Stolarski, et al., 2000) in about 12 A.D. Despite the early invention of the rolling element bearing it did not widely appear until the fifteenth century when Leonardo DA Venci (1452-1519 AD) had studied friction thoroughly and designed the first modern rolling elements bearing (Harris, et al., 2007). In 1794, a British Patent by Philip Vaughan for a radial ball bearing, which was considered the first of its type, appeared. The first mathematical formulation for modern rolling element bearings was developed by a German scientist Heinrich Hertz in 1881 (Stolarski, et al., 2000). The great expansion in using rolling elements bearings in industry started in the twentieth century, and particularly since 1960, after the great advancement in steel alloy production and high precision manufacturing making it possible to produce high precision and endurance rolling elements (Harris, et al., 2007). In 1966 the tribology field that is dedicated for friction, wear and lubrication studies was announced and has its name after the suggestion by Peter Jost (Popov, 2010). Thereafter, various designs and materials of rolling elements bearing come to life every year in response to new applications. Nevertheless, bearing failure has been one of the most frequently reported reasons for rotary systems breakdowns. It is therefore important to monitor bearing conditions and assess the Useful Remaining Life (URL) to avoid costly and sometimes catastrophic system failures. This obviously relies on the availability of reliable bearing fault detection and diagnostic methods.

2.2 Fault Detection and Diagnosis Methods for Rolling Element Bearings

Bearing fault detection and diagnosis were generally carried out using one or more of the following methods:

1. Acoustic emission analysis.
2. Vibration analysis.
3. Oil analysis.
4. Temperature analysis.

Considering space limit and pertinence to this thesis, only the vibration and temperature analysis methods are reviewed in this chapter.

2.2.1 Vibration analysis methods

The measured vibration signal from a mechanical system is often a mixture of several signal components from various sources. These signal components depending on their sources affect each other in different ways, e.g., summation, amplitude modulation or frequency modulation (Kurfess, et al., 2006). Extraction of bearing signal from this mixture is very important for early fault detection. However, this task is complicated due to the existence of noises that are different in nature including random noises with Gaussian distribution statistical behavior and deterministic noise such as that from gear meshing (Kurfess, et al., 2006). Many signal processing methods have been developed for bearing fault signature extraction. These methods can be categorized according to their objectives as follows:

1. Methods for random noise suppression.
2. Methods for separation of bearing signal from deterministic noise/signals.
3. Methods for strengthening bearing signals with respect to random or discrete noise.

Many techniques were used for bearing fault detection through accomplishing all or some of the above objectives. One of the earliest and most popular techniques for bearing signal detection is the High Frequency Resonance (HFR) method, also named envelope analysis or amplitude demodulation, which was proposed by Balderston (Balderston, 1969). This method was successfully used for monitoring gearboxes and turbo-machines as it was able to separate the signal of a defective bearing from the vibration signals from other system components (McFadden, et al., 1984). The reason that the envelope analysis method is effective for bearing faults detection is because all types of rolling element bearing experience some slip motion. Although this slip motion is in the order of 1-2% of the rolling motion, it does give a fundamental change in the character of the signal and that is why envelope analysis can extract diagnostic information that cannot be revealed by direct Fourier transform of the raw signal (Randall, 2011). However, the success of this method largely depends on the proper selection of signal passband and in-band signal enhancement. Band selection is to choose a center frequency and a bandwidth of a bandpass filter used in the amplitude demodulation method. In-band signal enhancement involves the application of denoising methods to the selected band to expose the target bearing signal. The tear down analysis (TDA) for some real cases of faulty gearboxes (Bechhoefer, et al., 2011) showed that improper selection of envelope band had resulted in gearbox complete failure before detecting any bearing faults. However, a slowly developing fault in a gearbox bearing can be detected by shifting the selected band by a few KHz. The very early technique for band selection was based on the experimental results, by examining the whole spectrum of the faulty bearing signal and comparing it with a healthy bearing spectrum.

2.2.1.1 Envelope window selection methods

The band which contains highest change due to the bearing fault is considered the best band for bearing envelope analysis (BEA) (McFadden, et al., 1984; Sawalhi, et al., 2004) and it is called Envelope Window (EW). In (Sheen, et al., 2012) the Linear Least Square (LLS) analysis was used for EW estimation. However, the method is largely affected by noise and for complex systems it becomes extremely difficult to use. Bozchalooi and Liang proposed a simpler method in (Bozchalooi, et al., 2007). This method uses a comparison between empirical Probability Distribution Functions (PDF) of the instantaneous frequencies for two data sets acquired in the same setting at two different speeds to detect the resonance frequency as the center of EW. Despite its simplicity, this method requires two conditions to be met. The first one is the prior knowledge of the fault frequency while the second condition is the assumption that the signal has only a single dominant resonance. Qiu *et al* applied a periodicity detection method for center frequency selection and used minimal Shannon entropy to choose the bandwidth (Qiu, et al., 2006). Lin *et al* (Qiu, et al., 2003) selected these parameters based on a kurtosis maximization criterion. Sawalhi and Randall (Antoni, et al., 2004) used the Spectral Kurtosis (SK) to select EW from the whole spectrum. Although SK was originally based on Short Time Fourier Transform (STFT), in this work they used both Wavelet Transform (WT) and Wigner-Ville transform (WVT) for bearing signal analysis. Their results showed similar results for the WT and the STFT analyses while the WVT suffered from some difficulties because of the anomalies due to window effect. The SK was also used in (Bechhoefer, et al., September 27-October 1, 2009) to select the best EW for helicopter gearbox bearing diagnosis. Bozchalooi and Liang (Bozchalooi, et al., 2008) proposed a smoothness index defined as the ratio of the geometric mean to the arithmetic mean of the wavelet coefficient moduli and used it to select the wavelet shape factor and to complement their work in the selection of the EW with the highest Signal to Noise Ratio (SNR). Despite the various methods proposed and tested for optimum EW selection, most of these methods suffer from either the high complexity or the lack of generality.

2.2.1.2 Filtering methods

The importance of vibration signal denoising is more obvious in the case of bearing fault detection and diagnosis because faulty bearing signal is more likely to be masked by signal components that have stronger vibration signatures (Sawalhi, et al., 2007). One of the oldest filtering methods is the time averaging. This method stems from the idea of removing the random noise that has a zero mean by time averaging the signal. This method was used for bearing fault detection by many researchers, e.g., Braun et al (Braun, et al., 1979) and McFadden (McFadden, et al., 2000) who used its variant, Synchronous Time Averaging (STA). In a recent review of most of averaging methods (Braun, 2011), Braun emphasized that revisiting old techniques like Time Domain Averaging (TDA) can reveal surprisingly that a new technique could have been presented and known for decades, but not with sufficient understanding.

Although the above filtering methods demonstrated their merits in removing random noise from the signal, they are not applicable to deal with dominant discrete noise that can mask bearing signal. Antoni et al (Antoni, et al., 2004) distinguished bearing signals from other discrete noise/signal using the concept of cyclostationarity. This concept was first proposed by Bennett in 1958, for communications applications, to describe random signals that are second orderly periodic (Gardner, et al., 1975). Based on this concept, many methods (e.g., (Antoni, et al., 2004; Antoni, et al., 2001)) were proposed to separate the bearing signal which can be considered quasi-periodic or Cyclo-Stationary (CS) from other deterministic signals like gears meshing signal.

2.2.2 Temperature monitoring methods

Temperature monitoring has been adopted for machine health assessment over the last few decades. Although most of temperature sensors are designed following similar physical principles, they differ in terms of complexity and precision. Thermography is one of the most advanced methods for temperature monitoring. Traditional systems that monitor the equilibrium temperature have found many applications including food and drug temperature monitoring. However, other applications such as high speed rotary machines need more sensitive health assessment systems that can detect faults in early stages. Therefore, instead of monitoring the equilibrium temperature, the temperature transient behavior for some applications is used for inspection and Non-Destructive Testing (NDT). Transient thermography, which employs pulse surface heating of an inspected component followed by acquisition of the thermal decay, was used for aircraft body inspection (Plotinkov, et al., 25-30 July, 1999). Thermography offers non-contact, wide area detection of subsurface defects, and can be used as an alternative or complement to conventional inspection technologies (Maldague, 2001). Passive thermography has been widely used in production, predictive maintenance, medicine, fire forest detection, building thermal efficiency survey, road traffic monitoring, agriculture and biology, medicine, gas detection and Non-Destructive Testing (NDT) (Maldague, 2001).

As mentioned above, rolling element bearings are widely used machine elements. Their failures are among the most frequently reported reasons for machine breakdowns. Recent studies on the application of thermography for bearing fault detection include those in (Kim, et al., 2012; Seo, et al., 2011). Researchers also worked on the development of wireless sensors for direct measuring bearing cage temperature (Marble, et al., 4-11 March, 2006; Scott, et al., 2011; Henao-Sepulveda, et al., 2005). The work related to modeling the thermo-mechanical behavior of rotating machines or the heat transfer mechanisms inside bearings has also been reported in the literature (Azad, et al., 2012; Yuzhong, 2006; Mizuta, et al., 2003). However, the potential of using temperature monitoring for the detection of both the physical bearing damages and lubrication problems has not yet been adequately addressed. The underlying cause that is hindering the use of thermal information for early fault detection for rotary machines in general and for bearings in particular is yet to be examined.

2.3 Friction in Rolling Contacts

The advances of rolling element bearing technology were mainly based on solving the problem of friction in rolling contacts. The efforts to explain and formulate this problem started in the 15th century when Leonardo Da Venci (1452-1519) recognized friction as a resistance to rolling motion. He stated that frictional forces were proportional to weight or normal force applied on rolling element and were independent of the area of contact. In 1699 Guillaume

Amontons rediscovered the frictional laws originally proposed by Leonardo da Vinci. Despite the acceptance of friction dependence on normal load, the idea of independence of friction on the area of contact was received with skepticism (Seireg, 1998). Charles Augustin Coulomb (1736–1806) is the most famous among the friction law founders. He recognized the importance of roughness and suggested that friction was due to the work done in dragging one surface up the other. One of the important contributions of Coulomb is his postulation that contact only occurs at the discrete points of asperity contacts. However, he rejected the adhesion theory and reasoned that if adhesion existed, the frictional resistance had to be doubled if the area of contact is doubled (Stolarski, et al., 2000).

2.3.1 Energy losses due to friction

For dry rolling friction there were several theories to explain the origin of energy loss during rolling. The most famous and earliest research was by Reynolds in 1875. He studied experimentally the dry fiction of rolling cylinder on a rubber surface. Reynolds concluded that the resistance to rolling motion was caused by the accompanying slippage. Later, both Palmgren in 1945 and Tabor in 1954, after repeating the Reynolds experiment in more detail, proposed that the interfacial slippage accompanying the rolling motion of an elastic material was very small and could not be the reason for the energy losses (Stolarski, et al., 2000). They proposed that the energy losses were mainly due to the hysteresis losses in the materials of both the rolling element and the surface (Seireg, 1998). The same result was emphasized by Greenwood and Tabor in 1958 (Greenwood, et al., 1958). Their experiment showed the importance of deformation losses as a main factor in specifying friction coefficient and consequently the accompanying energy losses.

In the recent literature of rolling element bearings, it is thought that the energy losses during bearing operation are the result of the combined effect of friction forces, elastic deformation hysteresis losses and drag force of lubricant film (Palmgren, 1959; Stolarski, et al., 2000).

2.3.2 Lubricant film thickness in Elasto-Hydrodynamic Lubrication (EHL)

Two important factors affecting energy losses in rolling element bearing are the properties of lubricant and its film thickness. In 1886 Euler was the first to define mathematically the lubricant pressure distribution and load-carrying capacity (Lugt, et al., 2011). In 1893 Barus mathematically described a relation between lubricant viscosity, pressure and temperature (Venner, et al., 2000). The first EHL was introduced in 1945 by Ertil (Cameron, 1985). His work established the definition of EHL that is still used now. Although the lubricant pressure distribution and film thickness were studied both experimentally and theoretically by many researches in the 50's and 60's of last century, the first reliable formula was reported as late as 1976 by Hamrock and Dowson (Hamrock, et al., 1976). The curve fit to their numerical solution is still the most widely used formula for finding film thickness in EHL. The next advancement in computing capabilities helped to solve the problem using multilevel techniques introduced by Lubrecht, et al. in 1987 (Lubrecht, et al., 1987). In 1994 Nijenbanning, et al. (Nijenbanning, et al., 1994) introduced a new formula for EHL film thickness based on the curve fit for several solutions using a MultiGrid technique (MG).

After the progress achieved in solving the EHL problem for fully flooded lubrication, researchers' interest was directed toward two other problems, i.e., the starved EHL and the rough surface. Damiens, et al. has studied both experimentally and numerically the effect of the operating conditions and lubricant properties on lubricant film starving (Damiens, et al., 2004). Their work showed that starvation severity depends on the type of contact, as well as the amount and thickness of the lubricant. After the invention of accurate, rapid shaping devices such as

Laser Surface Texturing (LST) and high-speed cameras, both micro-geometry surface shaping and accurate yet fast monitoring become possible. Tribologists exploited these recent technologies in designing experiments that was very difficult to be accomplished before. One of these advanced experiments was done by Mourier, et al. (Mourier, et al., 2006) to study the effect of circular micro-cavity passing through EHL contact. The results showed that this transient process can result in either increase or decrease in the lubricant film thickness depending on the relative dimensions of the cavity with respect to the contact dimensions. Recently, Boffy, et al. (Boffy, et al., 2012) presented an advanced 3D-model to solve the problem of moving heat source as a step prior to solving the 3-D thermo-elastic problem.

2.3.3 Friction reduction caused by vibration

In real life, many phenomena cannot be explained without considering the reduced friction caused by vibration. For example, steady objects on a table will start to move if the table vibrates. Also, a vibrating machine is more vulnerable to having loose fasteners in a shorter time than its non-vibrating counterparts. This phenomenon can be useful for some applications though detrimental to others. For example, vibration is used in packing industry to reduce the friction between the packaged items and the container.

Early attempts to study the effect of vibration on friction coefficient were made by Claypoole in 1943 (Godfrey, 1967) who found that the vibration can significantly reduce the static coefficient of friction. Fridman and Levesque in 1959 (Fridman, et al., 1959) studied the same effect of vibration on the static coefficient of friction. In 1967, both Godfrey (Godfrey, 1967) and Tolstoi (Tolstoi, 1967) investigated experimentally the effect of vibration on the kinetic friction coefficient. The work of Godfrey showed that the vibration in the vertical direction can reduce the kinetic coefficient of friction. Tolstoi et al. later modeled the contact region between two surfaces as a nonlinear spring, using an empirical stiffness relation (Tolstoi, et al., 1973). The theoretical understanding of this phenomenon was presented by Skamoto et al. (Sakamoto, et al., 1978; Sakamoto, 1987) and Hess and Soom (Hess, et al., 1991; Hess, et al., 1991). Their findings indicated that under harmonic loading the mean area of contact was decreased, leading to reduced mean friction force. Their model also suggested that the decrease in friction forces due to dynamic normal loading, when there was no loss of contact, was approximately 10% of the friction force due to static loading.

In a very recent application, vibration is used to control the surface friction of T-Pad application to create the sensation of the texture. In this application a piezoelectric bending moment is used to generate vibration of around 33 kHz with an amplitude of several microns (Winfield, et al., 2007). Also, in the medical field, it was proposed to vibrate the angioplasty device during insertion for friction force reduction (Sakamoto, et al., 2012). In the food industry, Schneider et al (Schneider, et al., 2009) found that vibrating the cutting tool made it easier to separate the material due to the decreased average sliding friction. They proposed that the temporary de-bonding of the contacting surfaces caused by ultrasonic vibration led to a reduction of the average friction.

Although the above studies have applied the vibration/oscillation to friction reduction in various fields, the exploitation of this phenomenon in bearing condition monitoring has not yet been reported in the accessible literature.

2.4 Summary

The review shows that many developments have been made in two separate fields, condition monitoring and tribology. However, an exploitation of the advancement achieved in these two fields together is rarely observed in the literature. Understanding the recent theories of tribology in addition to employment of recent temperature measuring devices can lead to new maintenance strategies that can be simpler and more indicative for mechanical systems than traditional strategies. The temperature rise in a rolling element bearing is one of the phenomena that have been studied by many tribologists. However the idea of relating temperature rise signature for a rotating equipment to its health condition and using this relation as a health assessment tool has not yet been reported in the accessible literature. On the contrary, temperature was always seen as a secondary tool in rotating machinery condition monitoring. This perception could change if a clear interpretation of temperature rise signature of rotating equipment during the thermally transient stage leads to a reliable health assessment indicator.

As mentioned above, vibration-based methods are commonly used for rolling bearing monitoring. Despite their successes, the vibration-based method are not suited to the detection of certain types of faults due to either the non-impulsive nature of some of these faults such as lubricant-related faults, or the time-varying nature of some other faults such as progressing spalls and cracks.

A system capable of both temperature and vibration monitoring should address the difficulties present in using each of the two monitoring methods alone due to their complementary features.

Chapter 3

Bearing Surface Temperature

The heat generated during bearing operation is from several sources: rolling friction, sliding friction, hysteresis losses, lubricant hydrodynamic resistance, rolling element-cage interaction and seal friction. Several studies were dedicated to find the factors affecting bearing heat generation sources and the accurate prediction of its rate. The heat generation rate and bearing surface temperature calculation at thermally transient state require two types of modeling that should be coupled to find out bearing surface temperatures. The first model solves for the bearing internal components interactions. This includes the calculation of forces, moments and relative speeds. The second model is used to describe the heat transfer network constructed from the bearing internal components thermal nodes.

Studies about measurement and calculation of rolling element bearing temperatures are among the most difficult problems in the fields of contact mechanics and measurement. The first reason of this difficulty is the complexity of the kinematics and dynamics of the bearing components. The second reason is the simultaneous occurrence of the complex combination of heat transfer modes in different media: solids, lubricant fluid and lubricant vapor and air. The third reason is the strong effect of lubrication that varies greatly due to changes of operating conditions, pressure and temperature.

The calculation of the rolling element bearing temperatures is done in two steps. The first step is to calculate the heat generated due to bearing operation. The second step is to solve the heat-transfer problem that controls the heat flow from/to each part of the bearing.

To calculate the bearing generated heat, all possible heat sources will be presented in section 3.1. Possible simplification for the radial ball bearing case is also described in the same section. Bearing heat generation calculation can be performed using empirical formulas or analytical models. Using empirical formulas for heat generation calculation will be shown in section 3.2. The usefulness of empirical methods is limited to the calculation of the bearing steady state temperature when operating conditions are well known and when there is no large deviation of bearing dimensions from the ideal design. Hence, these methods can be used to calculate the expected temperature for ideal bearing operation. A detailed presentation of analytic methods for bearing heat generation is shown in section 3.3. Although the analytical modeling is complex and requires extensive computational work, it is more accurate for calculating bearing transient temperatures. This is because the analytic methods are more flexible and can be used to model different factors affecting traction forces that generate heat.

The second step is to solve the heat transfer problem and can be done using analytical model or solving a finite element model for heat transfer using one of the commercially available finite-element software. In this study, the lumped mass method is used for bearing temperatures calculation. A brief description of the heat transfer model used to calculate bearing temperatures is shown in section 3.4.

3.1 Heat Sources in Rolling Element Bearings

Rolling element bearings, also called antifriction ball bearings, are commonly used to offer lower resistance to motion than sliding bearings do. However, the heat generated during an antifriction bearing operation is of great importance. Most of this heat is generated by friction moment caused by the resistance to rolling elements rotation. Although rolling friction or rolling resistance is of small magnitude, it is a complicated phenomenon and does contribute to the temperature rise, particularly in the presence of a fault. The accessible literature has not provided a theory for predicting the magnitude of the rolling resistance under all possible conditions of bearing operations due to the complexity of the issue (Stolarski, et al., 2000). This complexity is largely attributed to the different natures of friction sources in rolling element bearings. The generated friction moment represents an energy loss that is mostly converted into heat. The higher the generated friction moment, the higher the energy loss and hence the more heat will be generated during a bearing operation.

The total bearing friction moment consists of several constituents that can be considered as different sources different sources of heat energy. An equation to calculate the friction moment, encompassing all involved constituents, would be rather complex. A simple representation of this fiction moment is given as follows (Stolarski, et al., 2000):

$$M = M_s + M_{gyr} + M_{hys} + M_{de} + M_c + M_v + M_T \tag{3.1}$$

where M_s is the moment due to differential sliding, M_{gyr} is the moment due to gyroscopic effect, M_{hys} is the elastic hysteresis moment, M_{de} is the moment caused by the deviation of rolling elements from ideal design, M_c is the moment resulted from the sliding friction between the rollers and the edges of the raceways, M_v is the moment produced by shearing of the lubricant film and finally M_T is the moment due to the temperature effects (Stolarski, et al., 2000).

Equation (3.1) can be further simplified since several of its moment components are insignificant. For example, in the case of very small angles of contact (e.g., for radial ball bearings), the friction moment due to gyroscopic effect, i.e., M_{gyr}, can be neglected (Stolarski, et al., 2000; Harris, et al., 2007). The modern bearings are subjected to strict quality assurance with no or very little geometric deviation, and hence $M_{de}= 0$. Another simplification can be made is to consider that sliding occurs between the balls and the raceway only in the contact area. Therefore, there is no or very little contact between the balls and the raceway edges. Accordingly, one can assume $M_c = 0$. Furthermore, it is assumed that the temperature rise during bearing operation is not sufficient to cause changes in bearing internal dimensions, which leads to $M_T = 0$.

After the above simplifications the total friction moment can be re-written as:

$$M = M_s + M_{hys} + M_v \tag{3.2}$$

where M is the total generated friction moment, M_{v} is the generated moment when the load is zero, mainly due to the drag force caused by lubricant viscosity, M_{s} is the resistance to sliding motion and M_{hys} is the damping resistance caused by elastic hysteresis of the material (Stolarski, et al., 2000).

3.2 Empirical Formulas for Bearing Heat Generation Rate

Despite many empirical methods for bearing heat generation rate modeling were proposed several decades ago, there is a lack of agreement between the methodologies and the results of these methods, even when applied to the same case. This is mainly due to the large number of factors that affect the total heat generation rate. Each of these methods can be described to be more suitable for some types of bearings or some operating conditions.

Two of the most famous empirical formulas are presented in the next two sub-sections. The results of these two empirical methods will be used later to compare with the results of the proposed analytical model. The first method was proposed by Palmgren (Palmgren, 1959) while the second and more recent empirical method is the one proposed by SKF bearing company (SKF, 2003) in 2003.

3.2.1 Palmgren model

In this model, the total friction moment during the bearing operation is the sum of the load dependent moment, M_l, and the load-independent moment, M_v (Harris, et al., 2007):

$$M = M_v + M_l \tag{3.3}$$

The above two moments are calculated using the empirical formulas described by Equations (3.4) and (3.5) presented below (Harris, et al., 2007):

$$M_v = \begin{cases} 10^{-7} f_o \left(\nu_o\, n\right)^{2/3} d_m^3 & when\ \ \nu_o n \geq 2000 \\ 160 \times 10^{-7} f_o\, d_m^3 & when\ \nu_o n < 2000 \end{cases} \tag{3.4}$$

where n is the shaft speed in rpm, ν_0 is the lubricant kinematic viscosity under zero pressure in centistoke, d_m is the mean diameter of the bearing in millimeters and f_o is a factor depending on the type of bearing and the method of lubrication (Palmgren, 1959). The above equation shows that the moment due to lubricant viscous friction is constant for bearing operating at low speeds, i.e., when speed is lower than $2000/\nu_0$. At moderate to high speeds, the moment is proportional to $\left(\nu_0 n\right)^{\frac{2}{3}}$ and when the bearing is operating with constant speed, the moment is only proportional to $(\nu_o)^{2/3}$. M_l is given by:

$$M_l = f_1 F_\beta\, d_m \tag{3.5}$$

where F_β is the equivalent load that depends on the magnitude and direction of the applied load. It can be expressed as follows for radial ball bearing:

$$F_\beta = 0.9\, F_a \cot\alpha - 0.1\, F_r \qquad or \qquad F_\beta = F_r \tag{3.6}$$

where F_a is the axial load, F_r is the radial load and α is the angle of contact The equivalent load is chosen to be the larger of the two values calculated by Equation (3.6). For deep groove ball bearings for which the nominal value for the angle of contact α is zero, the equivalent load is given by:

$$F_\beta = 3\, F_a - 0.1\, F_r \tag{3.7}$$

f_1 in Equation (3.5) is a factor that depends on the bearing design and relative bearing load (Harris, et al., 2007):

$$f_1 = z\left(\frac{F_s}{C_s}\right)^y \tag{3.8}$$

where *z* and *y* are constants specified based on bearing type and design, F_s is the static equivalent load defined as the pure radial or pure axial load that causes the same permenant deformation at the heaviest loaded contact point as that of the applied combined loads, and C_s is the bearing static load rating. The basic load rating of a bearing is defined as the load applied to a non-rotating bearing that will result in a permanent deformation of 0.0001 *D* at the weaker of the inner or outer raceway contacts occurring at the position of the maximum loaded rolling element. *D* is the rolling element diameter. In the case of radial bearings that have a radial load, $\left(F_s = F_\beta = F_r\right)$ Then Equations (3.5) and (3.8) can be combined, resuling in a relation between the friction moment and the load, i.e.,

$$M_l = f_2 F^c \tag{3.9}$$

where f_2 is a constant depending on bearing size, construction and material, as well as the lubricant characteristic, F is the bearing operating load, and c is the exponent dependent on bearing type and design. It was found empirically that c varies between 1.0 and 1.2 for roller bearings with good roller guiding, and between 1.2 and 1.6 for various types of ball bearings under different load conditions. The value of c is essentially dependent on the sliding loss caused by the curvature in the contact surfaces (Stolarski, et al., 2000).

The rate of heat generation, H in watts, due to bearing operation can then be calculated by multiplying the total friction moment ($M_l + M_v$) given in Equations (3.4) and (3.9) by the shaft speed, n as follows:

$$H = 1.047 \times 10^{-4} (M_l + M_v) n \tag{3.10}$$

where M_l and M_v are the friction moments in N.mm and n is the shaft speed in revolutions per minute (rpm) and hence the resulting heat energy rate will be in watts.

3.2.2 SKF model

The model proposed by SKF bearing incorporation, is more recent and takes into account more factors to calculate bearing friction moment and resulted heat. The main advantage of this model over the others is the consideration of the lubricant change effect on all the friction constituents that build up the total friction moment. The total friction moment calculated by this model consists of four constituents, i.e. (SKF, 2003),

$$M = M_r + M_{sl} + M_{seal} + M_{drag} \tag{3.11}$$

where M is the total friction moment, M_r is the rolling friction moment, M_{sl} is the sliding friction moment, M_{seal} is the moment resulted from seal friction and M_{drag} is the friction moment resulted from the drag effect. The experimental work over a wide range of bearing designs and operating conditions leaded to the definitions of the moment components in Equation (3.11). This is described briefly as follows:

$$M_r = G_{rr} \left(\eta n\right)^{0.6} \tag{3.12}$$

$$M_{sl} = G_{sl} \mu_{sl} \tag{3.13}$$

$$M_{seal} = K_{s1} d_s^{\gamma_s} + K_{s2} \tag{3.14}$$

$$M_{drag} = V_M \ K_{ball} \ d_m^5 \ n^2 \tag{3.15}$$

where G_{rr}, G_{sl}, K_{s1}, K_{s2} , V_M and K_{ball} are constants depending on bearing design, loading condition and lubrication method (These constants can be found in the SKF bearing general catalogue (SKF, 2003)), d_m is the bearing mean diameter, n is the bearing speed in revolution per minute, d_s is the bearing shoulder diameter, γ_s is an exponent that depends on the type of the seal and μ_{sl}, is the sliding friction coefficient for full fluid film.

A detailed parametric study for bearing heat generation rates using both Palmgren and SKF models is described in chapter 4.

3.3 Analytical Solution for Bearing Heat Generation Rate

It was found experimentally that most of friction heat produced during rolling element operation is due to the sliding motion in the rolling contacts (Harris, et al., 2007).. The other friction constituent, rolling resistance, is so small compared to sliding friction such that it can be neglected. For lubricated bearing, the drag resistance of motion due to lubricant viscous effect is also a major source of bearing heat. The calculation of heat generation rate can be greatly simplified if some of the heat sources discussed in section 3.1 can be excluded. This simplification is decided with regard to each tribology system or the operating conditions. Hence, the analytic solution procedure for bearing heat generation during operation consists of the following steps:

1. Define all possible heat generation sources in the studied case. This step was presented in section 3.1.
2. Determine the load displacement relationship for internal parts (inner and outer races, rolling element and cage) and the relative speeds of contacting parts. This can be done based on the dynamic model of the bearing components.
3. Calculate relative velocities of bearing components.
4. Find the friction coefficient based on contact type (dry, boundary and hydrodynamic contact). This is done by entering the results obtained by the dynamic model into an Elastohydrodynamic (EHD) model to solve for the pressure distribution and the film thickness.

3.3.1 Speeds and forces for rolling bearing contacting elements

The calculation of the speeds of rolling element bearing components is essential to find the total friction and heat due to the relative motion of the bearing components. Although the kinematic solution (to be presented in section 3.3.1.1) for bearing component motions does not consider slippage, it is a first essential step to find out the orbital speed, n_m, and rolling speed of bearing rolling elements, n_R.

3.3.1.1 Kinematic solution for bearing speeds

The motions of internal components of a rolling element bearing are quite complex. These include the rotational motions of inner and/or outer rings, orbital motions of the rolling elements around bearing center and the spinning motion of each rolling element around its own axes. All these motions should be studied in order to calculate the sliding velocities or the relative velocities of contacting bodies. Firstly, the simple kinematic relations can be used to calculate the bearing cage speed. In most cases, the cage speed can be considered as the orbiting speed for the rolling elements. It is assumed that the relative motion between the rolling elements and the cage is negligible. In Figure 3.1, D_b is the ball diameter, D_i and D_o are the internal diameters of inner and outer races and d_m is the mean or the pitch diameter, n_o, n_i and n_R are the rotational speeds of outer race, inner race and ball respectively. The tangential speeds at ball center, ball contact point with inner and outer races are v_m, v_i and v_o respectively. Using simple kinematic relations, the orbiting speed of rolling element around bearing axis and the rolling speed of rolling element around its own axis can be obtained as follows:

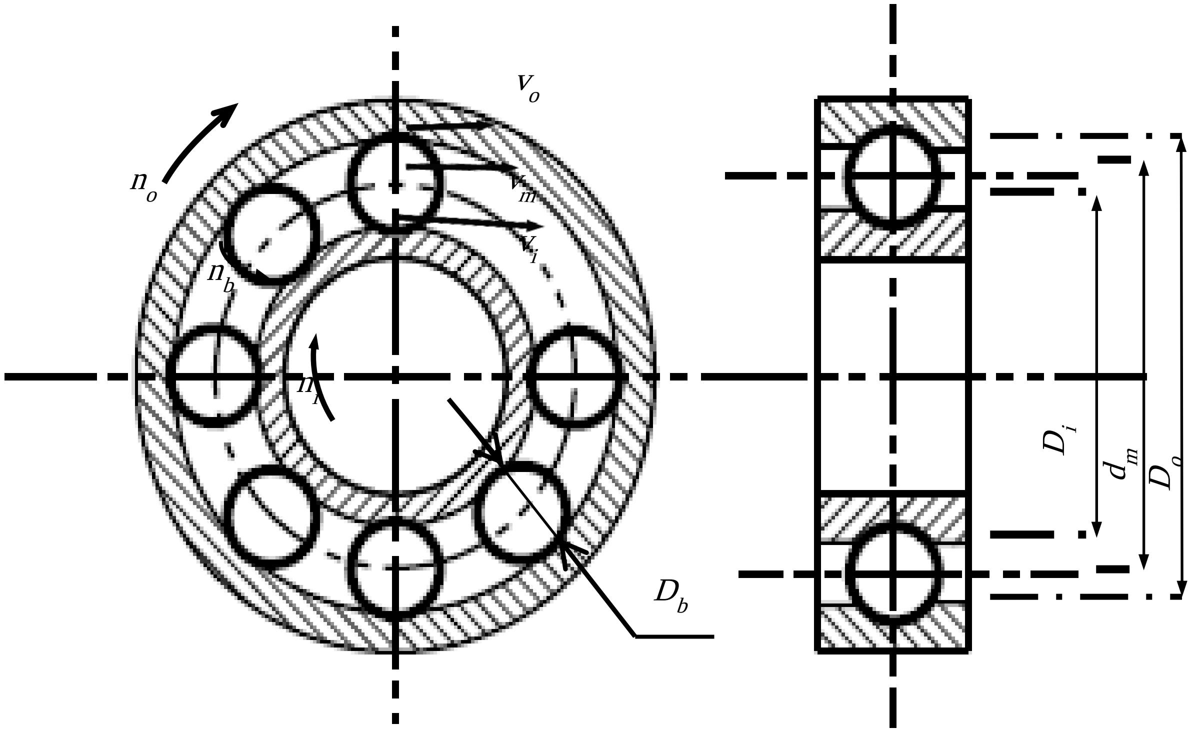


Figure 3.1 Radial ball bearing internal dimensions and speeds.

$$v_i = w_i \frac{D_i}{2} = \frac{1}{2} w_i \left(d_m - D_b \right) \tag{3.16}$$

where ω_i is the angular velocity of inner race given by:

$$\omega_i = 2\pi n_i \tag{3.17}$$

The tangential velocities at inner race, outer race and rolling element center are respectively calculated by:

$$v_i = \pi n_i d_m \left(1-\gamma\right) \tag{3.18}$$

$$v_o = \pi n_o D_m \left(1+\gamma\right) \tag{3.19}$$

$$v_m = \left(\frac{v_i + v_o}{2}\right) = \frac{\pi d_m}{2}\left(n_i \left(1-\gamma\right) + n_o \left(1+\gamma\right)\right) \tag{3.20}$$

where γ is equal to $\frac{D_b}{d_m}\cos\alpha$. The bearing rolling element orbiting speed n_m and the rolling element rolling speed n_R are given by (Harris, et al., 2007).

$$n_m = \frac{1}{2}\left(n_i \left(1-\gamma\right) + n_o \left(1+\gamma\right)\right) \tag{3.21}$$

$$n_R = \frac{d_m}{2D_b}\left(\gamma^2 - 1\right)\left(n_o - n_i\right) \tag{3.22}$$

From the above two equation the ball rotational angular speed, ω_R, and ball orbiting angular speed ω_m can be obtained as follows

$$\omega_R = 2\pi n_R \tag{3.23}$$

$$\omega_m = 2\pi n_m \tag{3.24}$$

The above two angular speeds will be used in the complete dynamic solution that will include the equilibrium equations for forces and moments, as discussed in the following section.

3.3.1.2 Complete dynamic solution

On contrary to the bearing kinematic calculation that assumes negligible slippage effect, discussed in the previous chapter, an accurate bearing heat calculation requires taking the slippage effect into consideration (Harris, et al., 2007). Although bearing kinematic calculations are not enough for finding the values of the generated heat and the bearing temperatures, it can be considered as the first step in finding the exact relative velocities of contacting bodies as a function of rolling angles that are performed by bearing rolling element. These angles and the resulting angular speeds for a rolling element in a radial ball bearing are depicted in Figure 3.2 as explained below. The figure shows the general case in which the axis of rolling of the rolling element is shifted by two angles: pivot angle β and spinning angle β'. On contrary to the pivotal motion that is resulted from the gyroscopic effect when the bearing angle of contact is not zero, The spinning motion of the rolling element is caused even without the need to the gyroscopic effect as a result of being the real contact area not a plane. In other words, the real area of contact has a curvature that has a harmonic mean profile radius. For point contact, such as the case of ball bearing, there are only two points that have a pure rolling motion. The rest of contact area points experience a combined rolling/sliding motion. In the next paragraph the detailed velocity components calculation is presented.

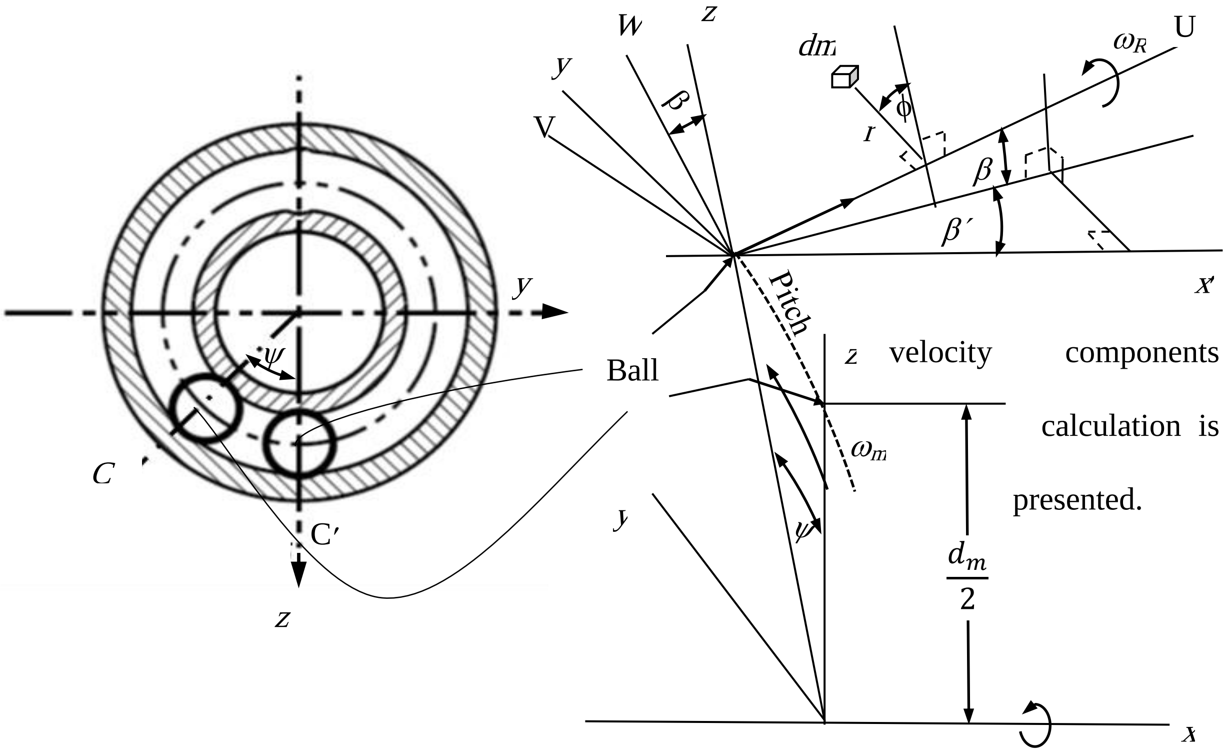


Figure 3.2 General motion descriptions for rolling element in a radial ball bearing.

In Figure 3.2, *Oxyz* is a Cartesian frame of reference that is fixed at the bearing center. In current case the outer race is assumed to be fixed, i.e., ω_i is the angular speed of the bearing inner race and the outer race angular speed ω_o=0. When the rolling element moves an angular distance, ψ, its center moves from point *C* to point C', *x'y'z'* is a Cartesian frame of reference that is fixed to the rolling element center and that is rotating about the bearing axis, *Ox*, with the orbiting angular velocity, ω_m .

When the rolling element can be considered completely constrained to the cage, the orbiting angular speed of the rolling element is the cage speed, which was previously calculated using kinematic Equations (3.18) – (3.20). Also, the rolling element has a rolling motion around its own axes, ω_R, which can be transformed to the orbiting Cartesian frame of reference, *x'y'z'*, in the form of three velocity components. Two of these velocity components are in the plane of the paper, $\omega_{x'}$ and $\omega_{z'}$, while the third component is the rotation motion around *y'* axis, $\omega_{y'}$. The angular velocities of the rolling element defined in *x'y'z'* frame of reference are as follows:

$$\begin{aligned} \omega_{x'} &= \omega_R \cos\beta \cos\beta' \\ \omega_{y'} &= \omega_R \cos\beta \sin\beta' \\ \omega_{z'} &= \omega_R \sin\beta \end{aligned} \tag{3.25}$$

where, β and β' are the angles of rolling about *C'y'* and *C'*z' axes, respectively.

The dynamic equations for a rolling element unit of mass d*m* are defined in a polar frame of reference, (*U*, *r*, φ). Depending on the above definitions and following the work done by Harris (Harris, et al., 2007) the sliding velocities at each of bearing races has two components, one in the bearing axial direction, *x* and the other in the bearing longitudinal direction, *y*.

These sliding velocities are given as a function of angles and curvature parameters at any point located at area of contact. Figure 3.3 shows the rolling element and the two areas of contact, A_i, the contact area between rolling element and the inner race, and A_o, the contact area between rolling element and the outer race. According to the Hertz theory, the contact area between a sphere and a flat surface is an ellipse. The dimensions of this contact area depend mainly on the radii of curvatures of the contacting bodies and the load. The areas of contact have semi-major and semi-minor axes, a_i and b_i for the inner race and a_o and b_o for the outer race. The lengths r_o and r_i are the radii of curvature for the outer and inner races grooves, respectively. R_i and R_o are the radii of curvatures for deformed surfaces at inner and outer races respectively and are given by the following equations respectively:

$$R_o = \frac{2r_o D_b}{2r_o + D_b} \tag{3.26 a}$$

$$R_i = \frac{2r_i D_b}{2r_i + D_b} \tag{3.26 b}$$

The dimensions of the elliptical areas of contact for steel bodies can be calculated using the following equations (Harris, et al., 2007):

$$a_j = 0.0236 a_j^* \left(\frac{3Q}{\sum \rho_j} \right)^{1/3} \tag{3.27}$$

$$b_j = 0.0236 b_j^* \left(\frac{3Q}{\sum \rho_j} \right)^{1/3} \tag{3.28}$$

$$\delta_j = 2.79 \times 10^{-4} \delta_j^* Q^{2/3} \left(\sum \rho_j \right)^{1/3} \tag{3.29}$$

In Equations (3.27)-(3.29), a_j and b_j are the semi-major and semi-minor lengths for the contact areas, j can be either i for inner race or o for outer race, Q is the applied normal force between the ball and the raceway, δ_j is the contact deformation in the direction perpendicular to contact region, a^*, b^* and δ^* are dimensionless parameters that can be either calculated using contact curvature parameters or the information given in tables that were produced using empirical formulas. These tables can be found in several references such as (Harris, et al., 2007) and (Brewe, et al., 1977).

The curvature sum $\Sigma\rho_j$ in the above equations depends on the curvature radii of the contacting bodies that are depicted in Figure 3.4 and is calculated using the following equation:

$$\sum \rho_j = \frac{1}{r_{I1}} + \frac{1}{r_{I2}} + \frac{1}{r_{II1}} + \frac{1}{r_{II2}} \tag{3.30}$$

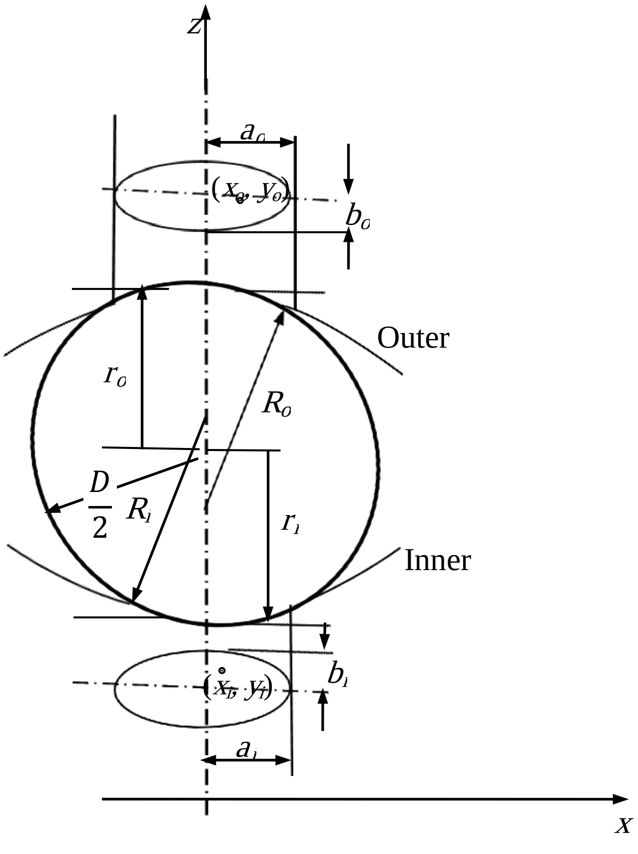


Figure 3.3 Radial ball bearing contact areas and radii of curvature.

In Figure 3.4, each of the two contacting bodies, I and II, has two different curvature radii in two perpendicular planes, plane 1 and plane 2. For example, the curvature radius $r_{\mathrm{II}1}$ is the radius of curvature of body II in plane 1 while $r_{\mathrm{I}2}$ is the radius of curvature of body I in plane 2.

After the calculation of contact area dimensions and defining the angular velocities of the rolling element in *x′y′z′* frame of reference as defined in Figure 3.2 and Equation (3.25), the expression for the sliding velocities at any point of the outer race contact (x_o,y_o) or inner race contact (x_i,y_i) and in the axial direction, *x* or the longitudinal direction, *y* can be given as follows:

$$v_{xo} = -\omega_{y'}\left[\sqrt{R_o^2 - x_o^2} - \sqrt{R_o^2 - a_o^2} + \sqrt{\left(\frac{D}{2}\right)^2 - a_o^2}\right] \tag{3.31}$$

$$v_{yo} = \omega_{x'}\left[\sqrt{R_o^2 - x_o^2} - \sqrt{R_o^2 - a_o^2} + \sqrt{\left(\frac{D}{2}\right)^2 - a_o^2}\right] \tag{3.32}$$

$$v_{xi} = -\omega_{y'}\left[\sqrt{R_i^2 - x_i^2} - \sqrt{R_i^2 - a_i^2} + \sqrt{\left(\frac{D}{2}\right)^2 - a_i^2}\right] \tag{3.33}$$

$$v_{yi} = -\frac{d_m\omega_i}{2} - \left[\sqrt{R_i^2 - x_i^2} - \sqrt{R_i^2 - a_i^2} + \sqrt{\left(\frac{D}{2}\right)^2 - a_i^2}\right] \times \left(\omega_{x'} - \omega_i\right) \tag{3.34}$$

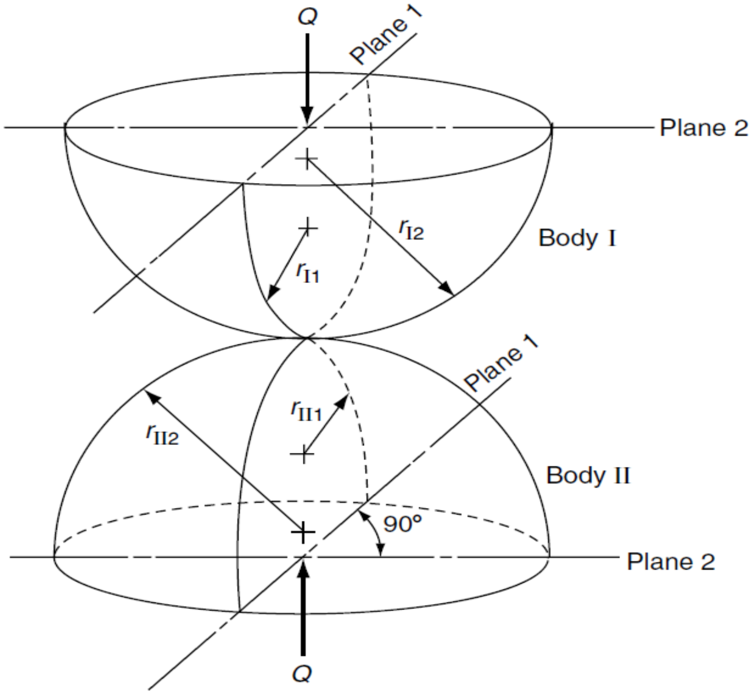


Figure 3.4 Curvature radii for two curved contacting hemisphere bodies.

In addition to the sliding velocities described in the four equations above, there is spinning motion which can be considered another source of friction. This spinning motion is resulted from the angular velocity component ωz' and it has the same value but opposite direction for inner and outer races. For radial ball bearing, the rolling element spinning motion at inner race ωsi and at outer race ω_{so} are given by

$$\omega_{si} = -\omega_{so} = \omega_R \sin\beta \tag{3.35}$$

The calculation of bearing sliding velocities at any point of the contact area using Equations (3.31) through (3.34) requires the calculation of the curvature radius at this point, Equation (3.26), the angular speeds at this point, Equation (3.25) and the contact area dimensions, Equations (3.28) and (3.29). The rolling element angles β and β' which are necessary to calculate angular sliding velocities, $\omega_{y'}$ and $\omega_{x'}$ in Equation (3.25) are calculated by analyzing the forces and the moments acting on every rolling element unit of mass, d*m* . The equations of Newton second law are considered for any unit of mass. in the polar coordinates (*U*, *φ*, *r*), shown in Figure 3.2, and then can be transformed to the general frame of reference *xyz* to solve for the rolling angles, β and β', through the following steps:

1. Transform the position vector for any unit of mass in the contact area from the polar coordinates (*U*, *φ*, *r*) to the ball-fixed frame of reference UVW, i.e.,

$$\begin{aligned} U &= U \\ V &= r\sin\varphi \\ W &= r\cos\varphi \end{aligned} \tag{3.36}$$

2. Transform the position vector from the ball-fixed frame of reference UVW to the orbiting frame of reference x′y′z′ using the following relations:

$$
\begin{aligned}
x' &= U\cos\beta\cos\beta' - V\sin\beta' - W\sin\beta\cos\beta' \\
y' &= U\cos\beta\sin\beta' + V\cos\beta' - W\sin\beta\sin\beta' \\
z' &= U\sin\beta + W\cos\beta
\end{aligned}
\tag{3.37}
$$

3. Transform the position vector from the bearing orbiting frame of reference x′y′z′ to the bearing inertial frame of reference xyz using the following relation:

$$
\begin{aligned}
x &= x' \\
y &= \frac{1}{2}d_m\sin\psi + y'\cos\psi + z'\sin\psi \\
z &= \frac{1}{2}d_m\cos\psi - y'\sin\psi + z'\cos\psi
\end{aligned}
\tag{3.38}
$$

4. Substitute Equation (3.36) into Equation (3.37) and then substitute the resulting expression for x', y' and z' into Equation (3.38). The result is an expression of the position vector for any bearing contact area unit of mass in the bearing fixed frame of reference, (x,y,z). The new expression defines the position vector in bearing inertial frame of reference as a function of the polar coordinates, (U, r, φ) and the bearing rolling angles, β and β' as shown below:

$$
\begin{aligned}
x &= U\cos\beta\cos\beta' - r\left(\sin\beta'\sin\varphi + \sin\beta\cos\beta'\cos\varphi\right) \\
y &= \frac{D_m}{2}\sin\psi + U\left(\cos\beta\sin\beta'\cos\psi + \sin\beta\sin\psi\right) + \\
&\quad r\left(\cos\beta\sin\varphi\cos\psi + \cos\beta\cos\varphi\sin\psi - \sin\beta\sin\beta'\cos\varphi\cos\psi\right) \\
z &= \frac{D_m}{2}\cos\psi + U\left(-\cos\beta\sin\beta'\sin\psi + \sin\beta\cos\psi\right) + \\
&\quad r\left(-\cos\beta'\sin\varphi\sin\psi + \cos\beta\cos\varphi\cos\psi + \sin\beta\sin\beta'\cos\varphi\cos\psi\right)
\end{aligned}
\tag{3.39}
$$

5. Apply the Newton’s second law, dynamic equations, for each unit of mass in the contact area while considering the position angles (ψ, β, β' and φ) as follows (Harris, et al., 2007):

$$
\begin{aligned}
\mathrm{d}F_x &= \ddot{x}\ \mathrm{d}m \\
\mathrm{d}F_y &= \ddot{y}\ \mathrm{d}m \\
\mathrm{d}F_z &= \ddot{z}\ \mathrm{d}m
\end{aligned}
\tag{3.40}
$$

and

$$\begin{aligned}\mathrm{d}M_{z'} = \Big\{ &-\ddot{x}\left[U\cos\beta\sin\beta' + r\left(\cos\beta'\sin\varphi - \sin\beta\sin\beta'\cos\varphi\right)\right] + \\ &\ddot{y}\left[U\cos\beta\cos\beta' - r\left(\sin\beta'\sin\varphi + \sin\beta\cos\beta'\cos\varphi\right)\right]\Big\}\mathrm{d}m\end{aligned} \tag{3.41}$$

$$\begin{aligned}\mathrm{d}M_{y'} = \Big\{ &\ddot{x}\left[U\sin\beta + r\cos\beta\cos\varphi\right] - \\ &\ddot{z}\left[U\cos\beta\cos\beta' - r\left(\sin\beta'\sin\varphi + \sin\beta\cos\beta'\cos\varphi\right)\right]\Big\}\mathrm{d}m\end{aligned} \tag{3.42}$$

where, F_x ,F_y and F_z are the total forces acting in *x*, *y* and *z* directions, respectively. Also, $\mathrm{d}M_{z'}$ and $\mathrm{d}M_{y'}$ are the incremental moments in z' and y' directions. It is assumed that the moment about *x* axis is zero as this is the condition for constant speed. Differentiating Equation (3.39) twice, the second derivatives of position vectors, ($\ddot{x}, \ddot{y},\ \ddot{z}$) can be expressed as a function of the mass element coordinate system, (*U*, *r*, φ) and the rolling element angular velocities, ω_R and ω_m as follows:

$$\begin{aligned}\ddot{x} &= r\omega_R^2\left(\sin\beta'\sin\varphi + \sin\beta\cos\beta'\cos\varphi\right)\\ \ddot{y} &= -2\omega_R\omega_m r\cos\beta\sin\varphi + \omega_m^2\left[-U\cos\beta\sin\beta' + r\left(-\cos\beta'\sin\varphi + \sin\beta\cos\varphi\sin\beta'\right)\right]\\ &\quad +\omega_R^2 r\left(-\cos\beta'\cos\varphi + \sin\beta\sin\beta'\sin\varphi\right)\\ \ddot{z} &= -2\omega_R\omega_m r\left(\cos\beta'\cos\varphi + \sin\beta\sin\varphi\sin\beta'\right) - \omega_m^2\left(\frac{D_m}{2} + U\sin\beta + r\cos\beta\cos\varphi\right)\\ &\quad -\omega_R^2 r\cos\beta\cos\varphi\end{aligned} \tag{3.43}$$

Substituting Equation (3.38) into the Newton Equations (3.40) to (3.42) yields four equations, two for force balance in *z* and *y* directions and two for moment balance in *z* and *y* directions with four unknowns, i.e., , *r*, β, β' and φ. The rolling element motions at any orbiting angle, ψ , can then be described by the solutions of the four equations.

The values of these angles at any point in the contact area will be used to find out the sliding velocities at each point and then these velocities will be used to calculate not only the friction torque in the dry contact but also the lubricant flow properties that will be used to calculate both the viscous drag friction and the hydrodynamic shear friction. This will be discussed in detail in the rest of this chapter and in the next chapter.

3.3.2 Surface friction shear stress in rolling element/raceway contact

The value of friction shear stress between rolling element and raceways depends on the type of contact, i.e., dry or lubricated contact. In the case of lubricated contact, the degree of contacting bodies' separation by the lubricant is a main factor to define the resulted shear stress due to rolling and sliding motion combination.

It is common to define three regimes of lubrication as described below:

1. Boundary lubrication -- the lubricant film thickness is very small and less than the surface roughness of the contacting bodies. The shear stress does not depend on lubricant viscous properties. The friction is determined by finding the properties of a boundary film that is a molecular blend of the contacting solid bodies and the lubricant.
2. Hydrodynamic lubrication or fluid film -- a complete separation of the contacting bodies is secured by a film of the lubricant. In this case, the friction shear stress is determined by calculating the viscous drag force that depends mainly on both the viscous properties of the lubricant and the operating conditions specified by load and velocity.
3. Partial lubrication -- this happens in two different situations: a) the lubricant quantity is not sufficient to completely separate the contacting bodies, and b) the applied load is very high and the operating speed is too low such that the fluid film of the

hydrodynamic lubrication contact can be penetrated and part of the load is carried by the direct contact of the contacting bodies asperities. This regime is also called mixed lubrication as it includes the simultaneous occurrence of both the boundary and hydrodynamic lubrication regimes.

Each of the above lubrication regimes has its own friction calculation methods that are shown briefly in the following subsections.

3.3.2.1 Boundary lubrication

When the lubricant film thickness is very small compared to surface roughness of contacting bodies, the most important parameters that govern the resulted friction are the normal stress distribution and friction coefficient. In this regime the friction coefficient can be considered constant and is independent on lubricant fluid viscosity (Hamrock, et al., 2004). The shape of contact area and the normal stress distribution can be calculated using Hertz theory. According to Hertz theory of elasticity the contact area for a sphere rolling on bearing raceway has an elliptical shape as depicted in Figure 3.4.

Consider the case of zero angle of contact (i.e., the radial ball bearing case) with a simple radial load Q. The maximum normal stress on the area of contact, σ_{max}, and the normal stress at any point in the area of contact, $\sigma_{(x,y)}$ are calculated as follows respectively (Harris, et al., 2007):

$$\sigma_{\max} = \frac{3Q}{2\pi ab} \tag{3.44}$$

$$\sigma_{(x,y)} = \frac{3Q}{2\pi ab}\left[1-\left(\frac{x}{a}\right)^2-\left(\frac{y}{b}\right)^2\right]^{1/2} \tag{3.45}$$

where a and b are respectively the semi-major and semi-minor axes for the elliptical area of contact, that are calculated previously using Equations (3.27) and (3.28). The shear stress at any point is then calculated based on the following relation:

$$\tau_{(x,y)} = \mu \cdot \sigma_{(x,y)} \tag{3.46}$$

where μ is the dynamic coefficient of friction that can be found for different tribology systems in many friction books such as (Seireg, 1998).

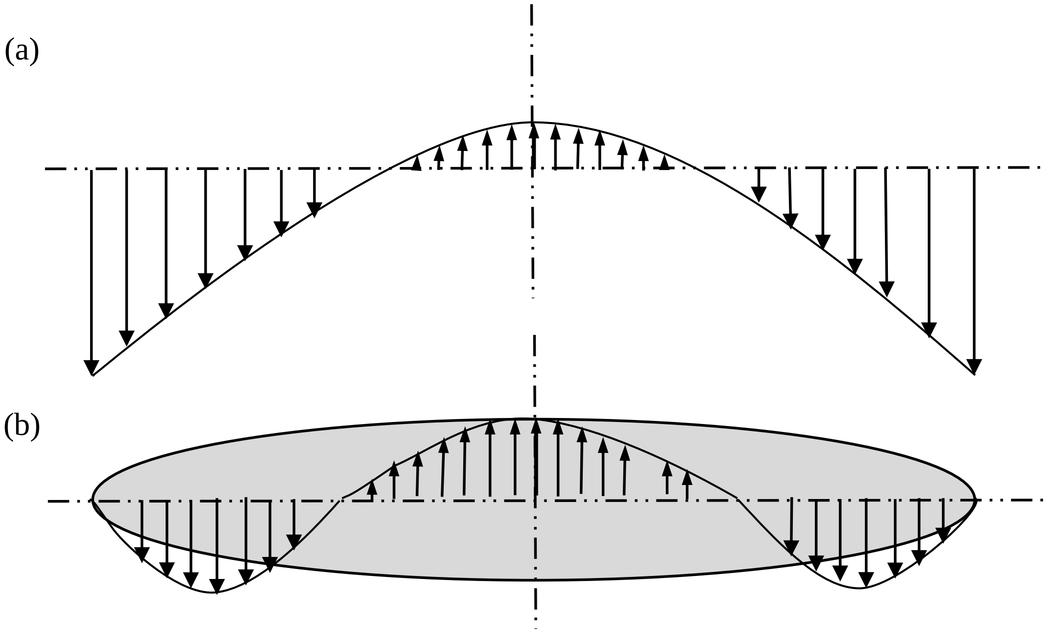


Figure 3.5 Distribution of: (a) sliding velocities, and (b) friction shear stress, τ, at any point, (x, y) of the elliptical area of contact (Harris, et al., 2007)

3.3.2.2. Hydrodynamic lubrication

In this regime the lubricant in the bearing plays a significant role by completely separating the rolling elements from the raceways, and then lubricant rheology is important to calculate traction. The shear stresses at any point in the contact region shown in Figure 3.6 should obey the following relations:

$$\tau_x = \eta \ \frac{\partial v_{xj}}{\partial z} \tag{3.47 a}$$

$$\tau_y = \eta \ \frac{\partial v_{yj}}{\partial z} \tag{3.47 b}$$

where τ_x and τ_y are the tangential shear stresses due to lubricant viscous effect in x and y directions, respectively, η is the lubricant viscosity, v is the tangential velocity of this point in the rolling contact that are given by Equations (3.26) through (3.29) and j can be either i for inner race or o for outer race.

As the gap distance between the contacting bodies is very small, Equations (3.47 a) and (3.47 b) can be replaced respectively by

$$\tau_x = \eta \frac{v_{xj}}{h} \tag{3.48 a}$$

$$\tau_y = \eta \frac{v_{yj}}{h} \tag{3.48 b}$$

where h is the lubricant film thickness that depends on the lubricant rheology. As the viscosity of the lubricant is not constant and varies with pressure and temperature, the above equation is only used for very limited cases of low loading conditions.

The viscosity of the lubricant is the most important factor when calculating traction forces and heat transfer rate. Although it is well known that, for almost all lubricants, viscosity increases exponentially with pressure and decreases exponentially with temperature, there is no equation that can relate the three parameters together for all lubricant types. However, in some ranges of operating pressure and temperature, there are some equations that can be useful to calculate lubricant viscosity at different pressures and temperatures. One of such formulas is the well-known Roelands equation (Harris, et al., 2007) given below:

$$\frac{\log_{10}\ \eta + 1.2}{\log_{10}\ \eta_0 + 1.2} = \left(\frac{T_0 + 135}{T + 135}\right)^{s_0} \left(1 + \frac{P}{2000}\right)^{z} \tag{3.49}$$

where η_0 is the viscosity and and T_0 is the temperature, in Kelvin degrees, at ambient conditions, η and T are respectively the viscosity and temperature at running condition, P is the applied pressure in kgf/cm^2 and s_0 and z are constants that need to be found experimentally for each lubricant type.

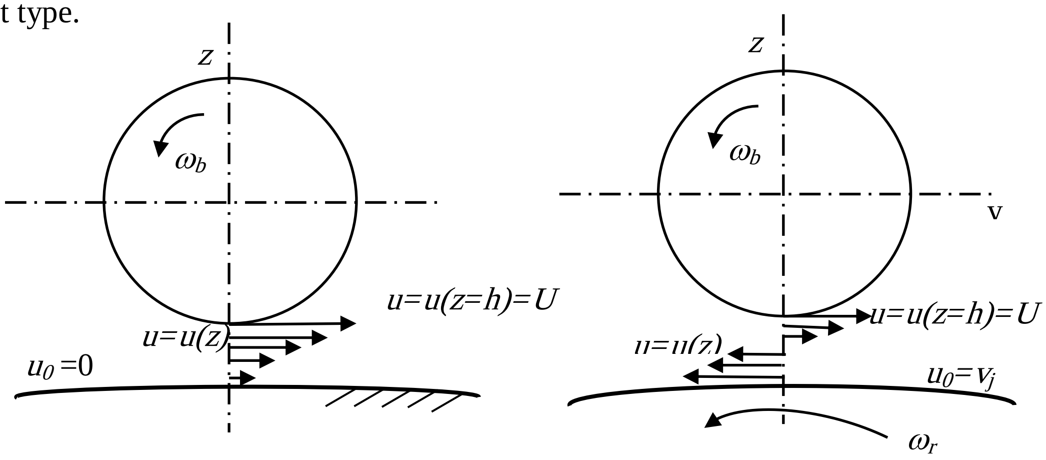


Figure 3.6 Lubricant tangential speed distributions for a) fixed raceway, and b) rotating raceways.

The modified lubricant viscosity calculated by the Reolands equation, Equation (3.49), will result in variable viscosity instead of the constant viscosity used in Equations (3.48 a) and (3.48 b). That is, the shear stress at different operating temperatures and loads can be expressed as follows:

$$\tau_x = \eta(P,T)\frac{v_{xj}}{h} \tag{3.50 a}$$

$$\tau_y = \eta(P,T)\frac{v_{yj}}{h} \tag{3.50 b}$$

The shear stresses calculated by the above equations are only valid for the Newtonian fluid assumption which can be only considered for low slide to roll ratios (<0.003), this shear stress can be defined as the Newtonian shear stress or τ_N. However, for higher values of slide to roll ratio, the maximum shear stress that can be sustained by the fluid, τ_{lim}, should be considered.

Considering the previous two definitions for Newotonian and limiting shear stress, the local contact friction shear stress can be calculated using following formula proposed by Trachman and Cheng (Trachman, et al., 1972):

$$\tau_f = \left(\frac{1}{\tau_N} + \frac{1}{\tau_{\lim}} \right)^{-1} \tag{3.51}$$

In the above equation, the friction shear stress τ_f is calculated for the given Newtonian shear stress τ_N calculated from Equations (3.50 a) and (3.50 b), and the limiting shear stress τ_{lim} that is usually found by the experimental testing. Some good experimental results by Schipper et al. (Schipper, et al., 1990) showed that the ratio of limiting shear stress to the average normal stress is in the range from 0.07 to 0.11.

3.3.2.3 Partial or mixed lubrication regime

In this lubrication regime both of the previous two regimes can occur in the same rolling contact. The calculation of traction forces in this regime requires the estimation of the ratio between the occurrence of boundary and hydrodynamic lubrication regimes, in other words, finding the ratio of the solid to solid contact area to the total contact area and using this ratio to calculate the total friction/traction shear stress. In this case there is a need to define the lubricant parameter Λ. This parameter is considered as a measure of the solid separation executed by the lubricant. When the lubricant parameter $\Lambda > 3$, a full separation is assumed and the formulas for full film lubricant can be used to calculate friction. For lower values of lubricant parameters, Λ, i.e. when $\Lambda < 1.5$, a boundary lubrication can be considered and Equations (3.44) through (3.46) are considered to calculate the friction shear stress.

In most cases the lubricant parameter, Λ, is in between the two limits, i.e., $1.5 < \Lambda < 3$.

Thence, the friction shear stress is calculated using the combination of the two equations, Equation (3.45) and Equation (3.50) as expressed by

$$\tau = c_v \frac{A_c}{A_o}\mu\sigma + \left(1-\frac{A_c}{A_o}\right)\left(\frac{1}{\tau_N}+\frac{1}{\tau_{\lim}}\right)^{-1} \tag{3.52}$$

In the above equation c_v is the velocity direction coefficient which can be either +1 or -1 for the two possible directions of sliding velocities as shown in Figure 3.6. Also, in Equation (3.52) the area of solid to solid contact is referred as A_c while A_o is the bearing apparent area of contact.

The total friction force acting on the whole area of contact is the result of integration of the shear stresses at all points of this area, i.e.

$$F_f = \int_{AC} \tau_{(x,y)}\ \mathrm{d}x\,\mathrm{d}\,y \tag{3.53}$$

where AC is the area of contact, F_f is the friction force in a tangential direction to the rolling direction. Combining Equations (3.45), (3.50) and (3.51), this friction force can be expressed using the following equation:

$$F_f = \int_{-b}^{b}\int_{-a}^{a}\left[c_v \frac{A_c}{A_o}\frac{3\mu Q}{2\pi ab}\left[1-\left(\frac{x}{a}\right)^2-\left(\frac{y}{b}\right)^2\right]^{1/2} + \left(1-\frac{A_c}{A_o}\right)\left(\frac{1}{\tau_N}+\frac{1}{\tau_{\lim}}\right)^{-1}\right]\ \mathrm{d}x\mathrm{d}y \tag{3.54}$$

It is clear from the above equation that to calculate the friction force, the ratio A_c/A_o should be first defined and calculated. This process for the lubricated types of contacts requires the calculation of the fluid film thickness and pressure distribution in the Elastohydrodynamic (EHD) contact. Several methods and models are previously represented for EHD point contact and some of these methods are discussed in the next section.

3.3.2.4 Lubricant film thickness and ratio for solid-solid contact

The lubricant forms a fluid film that has a significant role in preventing solid to solid contact and cause the pressure distribution in contact zone to differ from that calculated by the Hertz theory. The formation and the dimensions of the lubricant fluid film is governed by three types of physics relations, i.e., the fluid hydrodynamics governed by the Reynolds equation, the elasticity deformation governed by elasticity theory and the transformation of lubricant properties governed by the lubricant rheological properties (Stachowiak, et al., 2014). The minimum lubricant film thickness can be firstly found by the Hamrock and Dowson formula that is considered the first theoretical solution for smooth and fully flooded Elastohydrodynamic contacting surfaces (Hamrock, et al., 1977):

$$H_0 = \frac{3.63\bar{U}^{0.68}\mathcal{G}^{0.49}\left(1-e^{-0.68k}\right)}{\bar{Q}_z^{0.073}} \tag{3.55}$$

where $\mathcal{G}$ is a dimensionless material parameter, H_0 is a dimensionless film thickness, $H_0=h_0/R'_y$, R'_y is the equivalent radius in rolling direction, k is the ellipticity ratio= a/b where a and b are the semi-major and semi-minor lengths of elliptic area of contact, $\bar{Q}_z$ is a dimensionless load parameter and $\bar{U}$ is the dimensionless speed parameter. For a given material definition for both lubricant and contacting bodies the dimensionless material parameter, $\mathcal{G}$, can be calculated using the following equation:

$$\mathcal{G}=\lambda E' \tag{3.56}$$

where λ is the lubricant pressure-viscosity coefficient [m^2/N] and E' is the effective elasticity modulus of the two contacting bodies, *A* and *B*, which can be calculated from the following equation:

$$E' = \frac{2}{\left(\frac{1-\nu_A^2}{E_A} + \frac{1-\nu_B^2}{E_B}\right)} \tag{3.57}$$

The equivalent radius of curvature in the rolling direction, $R_y^{\backslash}$, for two bodies A and B is calculated as follows:

$$R'_y = \left(\frac{1}{R_A} + \frac{1}{R_B}\right)^{-1} \tag{3.58}$$

where ν_A, ν_B, E_A and E_B are respectively the Poisson ratios and the moduli of elasticity for the materials A and B. For example, the equivalent modulus of elasticity for two steel contacting bodies is $E' \approx 2.3\times10^{11}$ Pa.

In Equation (3.55) the other dimensionless parameters are defined as follows:

$$\bar{U} = \frac{\eta_0 U}{E' R'_y} \tag{3.59}$$

$$\bar{Q}_z = \frac{Q}{E' R'^2_y} \tag{3.60}$$

In the above two equations, U is the surface velocity, $U = (U_A+U_B)/2$, and Q is the normal applied load. Similarly, using the same dimensionless parameters, the plateau or central film thickness, H_{cen}, can be calculated using the following equation:

$$H_{cen} = \frac{2.69\bar{U}^{0.67}\mathcal{G}^{0.53}\left(1-0.61e^{-0.73k}\right)}{\bar{Q}_z^{0.067}} \tag{3.61}$$

As it can be seen from Equations (3.52) to (3.58) for lubricant film thickness calculation, the lubricant film thickness depends on lubricant properties in addition to the operating conditions. Both lubricant viscosity and lubricant pressure-viscosity coefficients play an important role in finding the lubricant film thickness. The methods to calculate the lubricant viscosity and the pressure viscosity coefficient are described in Appendix A.

To include the effect of the contacting bodies' surface roughness, a micro-contact model should be presented. One of the first and most popular models, the Greenwood and Williamson model or the GW model (Greenwood, et al., 1966) is used here to study the roughness effect on real area of contact. Although this model is based on several simplification assumptions, it showed efficiency in estimating the relative contact area A_c/A_0 when its results were compared with the results of experimental studies (Jeng, et al., 2003).

The GW model uses the contacting surfaces geometry represented by the parameters of surface roughness profile, R_s, D_{sum} and σ_s which are the radius of summits that are assumed to have a sphere shape, the density of summits per unit area and the standard deviation for summits height distribution that is a normal distribution, respectively. In addition to the above three parameters for the micro-contact, there are three spectral moments for the summit heights distribution, i.e., the zeroth, second and fourth moments of profile distribution. These three parameters are denoted m_0, m_2 and m_4 and defined as follows respectively:

$$\begin{aligned} m_0 &= E\left(z^2\right) = S^2 \\ m_2 &= E\left[\left(\frac{\mathrm{d}\,z}{\mathrm{d}\,x}\right)^2\right] \\ m_4 &= E\left[\left(\frac{\mathrm{d}^2\,z}{\mathrm{d}\,x^2}\right)^2\right] \end{aligned} \tag{3.62}$$

where $z(x)$ is the surface profile in direction x which can be measured experimentally in addition to its slope that is equivalent to $\left(m_2\right)^{\frac{1}{2}}$. The fourth moment cannot be measured and is usually calculated with computer codes and E stands for the probability expectations operator (Harris, et al., 2007).

The steps for using the GW model are as follows:

1. Calculate the composite surface roughness, S, of the two bodies, shown in Figure 3.7(a), that have a rms surface roughness S_A and S_B by,.

$$S = \sqrt{S_A^2 + S_B^2} \tag{3.63}$$

2. Use the profile distribution parameters m_0, m_2 and m_4 to calculate asperity density, D_{sum} , the bandwidth parameter, α_{bw} and the lubricant parameter, Λ, as described below.

$$D_{sum} = \frac{m_4}{6\pi m_2 \sqrt{3}} \tag{3.64}$$

$$\alpha_{bw} = \frac{m_0 m_4}{m_2^2} \tag{3.65}$$

$$\Lambda = \frac{H_{cen}}{\sqrt{m_0}} \tag{3.66}$$

3. The distance between the summit mean plane and surface mean plane, $\bar{Z}_s$,shown in Figure 3.7(b), is calculated using the modified GW model proposed by Bush et al. (Bush, et al., 1975):

$$\bar{Z}_s = \frac{4S}{\sqrt{\pi \alpha_{bw}}} \tag{3.67}$$

4. The probability of solid to solid contact is the same as the probability of having an asperity that has a height Z_s greater than distance d. Here d is the maximum height that will not cause an asperity contact and can be related to the lubricant film thickness that is calculated previously using Equation (3.61). Hence

$$d = H_{cen} - \bar{Z}_s \tag{3.68}$$

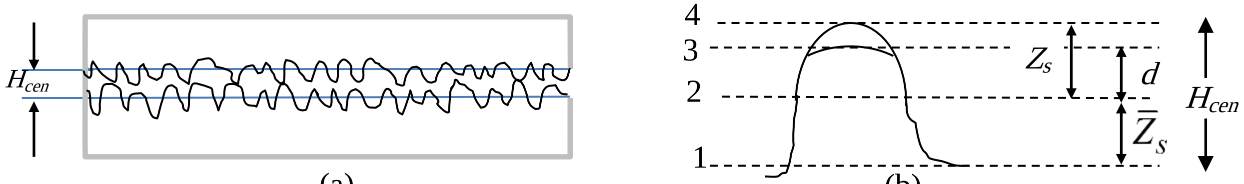
5. The standard deviation for the surface roughness was defined by Bush et al. (Bush, et al., 1975) as follows:

$$S_s^2 = \left(1 - \frac{0.8968}{\alpha_{bw}}\right) S^2 \tag{3.69}$$

6. Substituting Equations (3.67) and (3.69) into Equation (3.68), the ratio of the asperity height to its standard deviation can be established as a function of the lubricant film thickness. That is

$$\frac{d}{S_s} = \frac{H_{cen} - \bar{Z}_s}{\sqrt{1-\frac{0.8968}{\alpha_{bw}}}\,S} = \frac{\left(H_{cen}/S\right) - \left(4/\sqrt{\pi\alpha_{bw}}\right)}{\sqrt{1-\frac{0.8968}{\alpha_{bw}}}} \tag{3.70}$$

7. The resulted ratio from Equation (3.70) is used to calculate the number of summits that cause asperity contact by looking in the probability tables for normal distribution. As the real area of contact and the load carried by asperities are functions of the number of asperities that contact, the real are of contact, A_c and the ratio A_c/A_0 can be calculated depending on the area under the normal distribution curve for the function that describes the relation of these values to the summits' height.



(a) (b)

Figure 3.7. a) The mixed lubrication between two rough surfaces, and b) details of a single asperity showing different planes: plane (1) is the mean plane for the rough surface, plane (2) is the mean plane for the summits, plane (3) is the largest height before possible solid to solid contact and plane (4) is the summits height plane

3.3.3 Viscous drag force

In addition to the friction resulted from shear stress in the rolling contact discussed in the previous section, the lubricant drag effect is taken into consideration in the current section. The viscous drag force between the ball and the lubricant, $F_{v,}$ in a ball bearing is given by (Harris , et al., 1973) and cited below:

$$F_v = \frac{c_d\ \pi D_b^2\ \xi\left(D_m\omega_m\right)^{1.95}}{32g} \tag{3.71}$$

where D_b is the ball diameter, D_m is the bearing mean diameter, ω_m is the angular speed at the mean diameter, c_d is the viscous drag coefficient that can be obtained from (Streeter, 1962) and ξ is the ratio of the weight of the lubricant in the bearing free space to the volume of the free space.

3.3.4 Total heat generation rate

The total heat generated due to bearing rolling and sliding friction can be obtained by adding the heat generation rate resulted from different types of friction sources. In the ball bearing case, these sources are the friction in ball-raceway contact (RBC) friction and the friction due to lubricant viscous drag. The heat generation rate due to RBC is calculated by the integration of the friction torque multiplied with the rolling speed over the area of contact. Therefore, for each rolling element, the total heat generated by rolling and sliding friction in y direction can be expressed using Equation (3.69) as shown below:

$$H_{ynj} = \int \tau_{ynj} n_{ynj} \mathrm{d}A_{nj} = a_{nj} b_{nj} \int_{-1}^{1} \int_{-\sqrt{1-q^2}}^{\sqrt{1-q^2}} \tau_{ynj} \nu_{ynj} \mathrm{d}t \mathrm{d}q \tag{3.72}$$

where H_{ynj} is the heat generated in N.m, τ is the shear stress, A is the area of contact and n or ν is the sliding velocity. The subscripts are defined as follows: j is the ball number (from 1 to z where z is the number of the balls), y is the direction of the calculated heat component and n is an indicator of inner or outer race (n = 1 for inner race and 2 for outer race). Variables t and q are limiting parameters for the integration and defined as $t = \frac{x}{a}$, $q = \frac{y}{b}$ where a and b are the semi-major and semi-minor axes of the elliptical area of contact. A similar equation can be used for calculation of the heat generated for each rolling element due to sliding in *x* direction:

$$H_{xnj} = \int \tau_{xnj} n_{xnj} \mathrm{d}A_{nj} = a_{nj} b_{nj} \int_{-1}^{1} \int_{-\sqrt{1-q^2}}^{\sqrt{1-q^2}} \tau_{xnj} \nu_{xnj} \mathrm{d}t\mathrm{d}q \tag{3.73}$$

Then the total heat generated by RBC for all the rolling elements at both outer and inner races is expressed as follows:

$$H_{RBC} = \sum_{n=i}^{n=o} \sum_{j=1}^{j=z} \left(H_{ynj} + H_{xnj} \right) \tag{3.74}$$

The heat generated due to lubricant viscous drag effect is given by

$$H_{drag} = \frac{D_m w_m F_v \ \mathrm{Z}}{2} \tag{3.75}$$

where Z is the number of rolling elements, and F_v is the viscous drag force as given by Equation (3.71). The total heat generation rate is the sum of the heat rates by Equations (3.74) and (3.75). as shown in the next t equation. Figure 3.8 shows the scheme for total friction calculation.

$$H_{tot} = H_{RBC} + H_{drag} \tag{3.76}$$

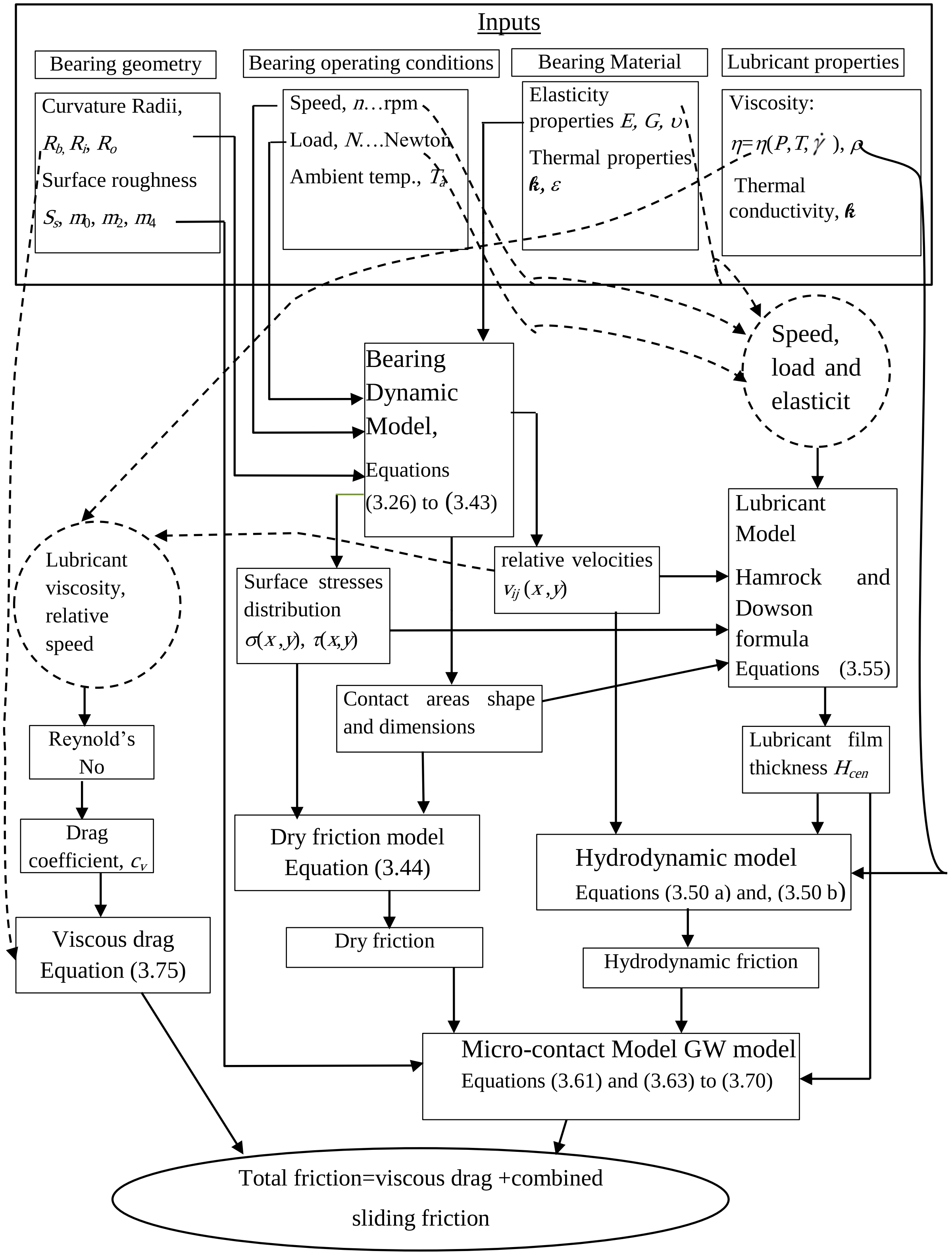


Figure 3.8 Calculation scheme for the total friction for lubricated rolling element bearing.

3.4 Rolling Element Bearing Temperatures and Heat Transfer Modeling

The bearing temperature is governed by two opposite processes, heat generation due to RBC friction and lubricant viscous drag and heat dissipation due to the structure and the surrounding medium. These two processes can run in one, two or all of three different modes of heat transfer involved in any heat transfer process.

In many tribology systems/subsystems only one or two heat transfer modes can be considered as the dominating modes and the other mode/modes can be neglected. The consideration of the dominating heat transfer mode depends on the process nature and the measuring system. In the case of bearing temperature measurement, the dominating mode depends mainly on both the measurement spot and the measurement device/sensor. For example, when measuring outer race temperature using a thermocouple, both conduction and convection modes are considered. The conduction is between rolling elements and the outer race as heat in-flow whereas the conduction from the outer race to the housing as a heat out-flow. On the other hand, the convection mode is mainly because of the heat transferred to the outer race via the lubricant film in the vicinity between rolling elements and inner wall of the outer race.

If thermography is adopted to measure the temperature of a rotating inner race as the highest temperature spot, radiation is the only mode. The radiating heat transfer from the inner race surface monitored by a thermal camera depends on the temperature difference between the inner race surface and the camera sensor, the distance, the bearing ring surface material emissivity and the ambient temperature of the medium through which the heat is transferred.

The heat calculated by Equation (3.76) is the total heat generated in one second and can be used to calculate bearing temperatures. For each bearing internal component *i*, the change in its internal energy $\Delta\mathscr{P}$ is the net result of the heat transferred to it from other components with higher temperatures and the heat dissipated from the component to other components that have lower temperatures. Then the total change in the component internal energy, $\Delta\mathscr{P}_i$, neglecting mechanical work exerted by the friction moments, can be expressed as:

$$\Delta\mathscr{P}_i = \sum_{j=1}^{n} H_{ji} - \sum_{k=1}^{m} H_{ik} \tag{3.77}$$

where H_{ji} is the heat energy transferred to component i, from components with higher temperatures j, (j=1…n). H_{ik} is the heat dissipated from component i to other components k, (k=1…m) that have lower temperatures. Knowing that the heat transfer to or from any component of the bearing is the result of three different modes of heat transfer, namely, conduction, convection and radiation, then the above equation becomes:

$$\Delta\mathscr{P}_i = \sum_{j=1}^{n}\Big(\big(H_{ji}\big)_C + \big(H_{ji}\big)_V + \big(H_{ji}\big)_R\Big) - \sum_{k=1}^{m}\Big(\big(H_{ik}\big)_C + \big(H_{ik}\big)_V + \big(H_{ik}\big)_R\Big) \tag{3.78}$$

where C, V and R stand for conduction, convection and radiation, respectively. Then the change in internal energy for each component i can be used to calculate the component temperature change using the following formula (Lienhard IV, et al., 2003):

$$\Delta T_i = \frac{\Delta\mathscr{P}_i}{m_i \boldsymbol{c}} \tag{3.79}$$

where, ΔT_i is the temperature change of bearing component i , m_i is its mass and $\boldsymbol{c}$ is the specific heat capacity for the material of component i.

For the calculation of the rate of change of internal energy for each point in the bearing internal components, a detailed thermal network should be constructed. The major points that have significant heat flow rates are defined on this network will be shown in next chapter.

Chapter 4

Calculation of Bearing Temperature

Usually the bearing temperature is calculated at the thermal equilibrium condition. This happens when the generated heat due to bearing operation and the total dissipated heat, by different heat transfer modes, are balanced. The thermal equilibrium condition for rolling element bearing usually takes very long time to reach. For many critical applications, the development of a fault and the occurrence of system failure can occur even before reaching the thermal equilibrium condition.

In this chapter an analytical model based on the theory presented in the previous chapter is proposed to calculate the temperatures of the bearing surface and internal components at the transient state, i.e., before reaching thermal equilibrium condition. The results obtained from this model are compared to the results by two of the popular friction models: Palmgren and SKF models, and to experimental measurements. The experimental measurements in the current work are done using several temperature monitoring devices, i.e., a thermal camera, an infrared sensor and the K-type thermocouples. The bearing transient temperature model proposed in current chapter will be used later in Chapter 6 to predict the bearing fault effect on bearing transient temperature curves after adjusting the model for different fault types effect.

4.1 Bearing Temperature Model

The task for modeling and calculation of bearing surface temperature in thermal transient state is complex requiring an integrative model that consists of three modules. The first module will calculate the stresses' distribution and the speeds of the points in the contact area considering the simplified assumption of the controlling race.

The second module will use the speeds and stresses calculated by the first module to define the lubricant flow characteristics. The module then use these characteristics to solve for the lubricant film thickness, Reynolds and Prandtl numbers, drag coefficient and lubricant thermal properties. The previous data are used to calculate both the hydrodynamic friction and the viscous. At this moment the main program use the calculated frictions by the first and second module to calculate total heat generation rate.

Then the third module uses the heat transfer equations to calculate temperatures at different bearing internal parts. The change of the lubricant properties due to the temperature increase will be taken into account and will be used to repeat the calculations of first and second modules until converging lubricant properties are attained. Each of the three modules will work as an independent module or independent Matlab function. However, there is a main program that controls the transfer from one module to another and does some correction iterations until reaching a converging solution.

4.1.1 Bearing dry friction calculation -- Module 1

The first module can be considered as a simplified dynamic model for bearing operation as it uses the controlling race assumption to simplify the dynamic equations shown in previous chapter (Equations (3.25) to (3.43)). The controlling race assumption is an effective method to calculate the sliding angular velocities in the rolling element bearing. It was proposed by Jones (Jones, 1959) and then re-applied and verified by Hamrock (Hamrock, 1975) for the arched outer race ball bearing and for different types of rolling element bearing by Harris (Harris , et al., 1973; Harris, 1971). In this assumption the sliding is considered one of the two bearing races while the motion of the ball on the other race is assumed to be pure rolling. Using this assumption the sliding angle in the rolling direction, *β*, is calculated by the following equation:

$$\tan\beta = \frac{\frac{1}{2}d_m * \sin\alpha}{\frac{1}{2}d_m \cos\alpha + r'} \tag{4.1}$$

where, d_m is the mean bearing diameter, α is the contact angle for the controlled race and r' is the radius of pure rolling. For calculation of the sliding angle, *β*, using the above equation the radius of pure rolling is assumed to be equal to half of the ball diameter and the angle of contact is given a very small value, $\alpha = 0.01$. According to this simplification, the angular velocities of ball bearing and the resulted sliding velocity at each point of the contact area can be calculated.

For the calculation of friction forces at each point in the contact area and based on the idea of the boundary element method (BEM), the contact area pressure distribution, deflection and friction force are calculated. Then by integrating the friction force over the complete area of contact, the total friction force and the heat generation rate can be calculated. The BEM method is suitable for the contact problems as it simplifies the problem by only discretizing the surface of the contact area (Popov, 2010). The other advantages of this method is the good accuracy and the adaptation to rapid changes of the studied boundary stresses (Becker, 1992). Instead of considering a continuous area of contact, the contact area is divided into a number of discrete surface elements, i.e., 2D segments. The normal pressure distribution for Hertzian contact of two elastic bodies in half space is considered as the friction stress is related to the normal stress by defining the coefficient of friction for the rolling contact between steels. In the above calculation of the normal and tangential stress distributions, the dynamics equations (Equations (3.42) to (3.45)) described in chapter 3 are applied.

The boundary conditions for the current problem are defined in two different types of regions that are defined in the contact area, the stick zone and the slip zone. The result of solving Equations (3.26) to (3.29) will define these two zones that were explained by Johnson (Johnson, 1985) as follows:

1. For the stick zone, the relative velocities at each point are the same at the two bodies, i.e. the total strain at any point, (x_0 ,y_0) of the stick zone, $\dot{S}_{(x_0,y_0)}$ equals zero. Also the traction component at each point cannot exceed the friction limit or,
   $$\left|\tau_{(x_0,y_0)}\right| \leq \mu\sigma_{(x_0,y_0)} \tag{4.2}$$
2. For the slip zone, the strain at any point can take any value and the friction shear stress equals the friction limit but in opposite direction, i.e.,
   $$\left|\tau_{(x_0,y_0)}\right| = \mu\sigma_{(x_0,y_0)} \tag{4.3}$$
3. The pressure at the edge of the contact area vanishes.

Considering the above boundary conditions and applying the controlling race assumption, relative speeds and accelerations are calculated at each point in the contact areas formed by the rolling element and both inner and outer races. The body forces resulting from the rolling elements inertial forces will be neglected except the centrifugal forces. The centrifugal forces generated at each azimuth angle will be added to the external radial force component in normal direction.

Assuming that the bearing temperature rise will not remain for a long time and will not reach high temperatures, the slight change in bearing internal dimensions will not be considered in current calculations. The Hertzian solution for the contact area dimensions and normal pressure distribution is considered as a primary solution. Then the module divides each contact area into 80 small areas, 10 divisions in the major axis and 8 divisions in the minor axis. The dynamic equations are solved at each point of the contact area. For the area divisions the load sharing and deflection is calculated separately from other points after the solution is modified by changing the forces affecting each rolling element. The current module will calculate the total friction shear stress considering dry contact and will be held to be used in the total shear stress calculation formula based on Equation (3.52).

The normal stress distribution at a contact area is depicted in Figure 4.1. This figure shows the contact area dimensions in *x* and *y* directions while the *z* direction is associated with the normal stress which is a compressive stress that reaches maximum at the center and minimum at the four corners of the contact area. The friction force at each point can be assumed in a direction tangential to the rolling direction and can be simply calculated assuming dry contact and using the friction coefficient for the steel-steel contact tabulated in many references, e. g. (Seireg, 1998; Tea, et al., 2007)

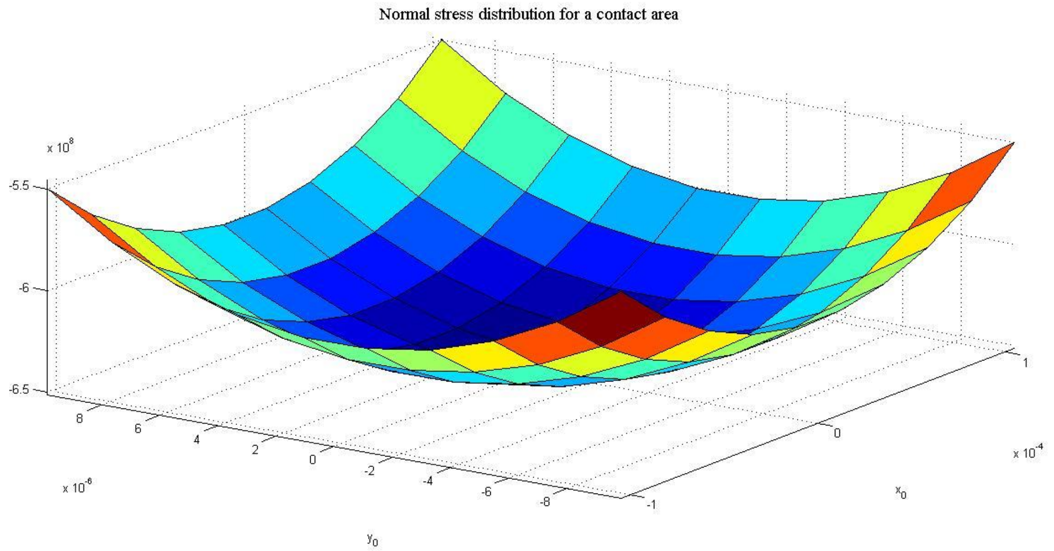


Figure 4.1 Normal stress distribution in a bearing rolling contact area

4.1.2 Bearing hydrodynamic friction calculation – Module 2

The lubricant effect on friction and consequently on the heat generation rate can be studied by dividing it into two separate parts. The first part is the friction generated due to the sliding between moving parts and lubricant. When the applied load is small and the operating speed is high, the full fluid film lubrication regime can be assumed. The full fluid film lubrication regime assumes that the lubricant can completely separate the two contacting bodies. The shear stress resulted from the relative motion of bearing race/rolling element with respect to the lubricant can be calculated by solving the Reynolds's equation taking into consideration possible changes of lubricant rheology due to the change of pressure, temperature and shear rate.

Considering a full fluid film and using Equations (3.50a) and (3.50b), the friction shear stress is calculated in two directions. For the steady state operation of the ball bearing, the lubricant film thickness can be calculated using the Hamrock and Dowson formula (Hamrock, et al., 1976) as in Equation (3.61) for the central fluid film thickness. The study of the property change of the lubricant or the lubricant rheology is reported in Appendix A while the speeds at each point will be extracted from the results of the previous module.

The second part of this module is to calculate the drag friction that is described by Equation (3.71). The friction force resulted from drag effect depends mainly on the coefficient of drag, c_d, the relative quantity of the lubricant in the bearing internal space, ξ and the orbiting angular velocity, ω_m. The drag coefficient for the flow of the lubricant relative to the bearing parts is given for three different flow regimes, i.e., laminar, transition and turbulent. The ranges for each flow condition are defined by a Reynolds's number, *Re*, i.e., range as follows (Streeter, 1962):

$$c_d = \begin{cases} \dfrac{1.328}{\sqrt{\mathrm{Re}}} & \mathrm{Re} < 5\times10^5 \\ \dfrac{0.074}{(\mathrm{Re})^{1/5}} - \dfrac{1700}{\mathrm{Re}} & 5\times10^5 < \mathrm{Re} < 10^7 \\ \dfrac{0.074}{(\mathrm{Re})^{1/5}} & 10^7 < \mathrm{Re} \end{cases} \tag{4.4}$$

The Reynolds's number is calculated at each area of contact using the average sliding velocity and the characteristic length, taken as the contact area length in rolling direction. The Reynolds's number is defined as follows:

$$\mathrm{Re} = \frac{\rho v_m l}{\eta} \tag{4.5}$$

where ρ is the lubricant density, η is the lubricant dynamic viscosity, v_m is the average value of orbiting speed and l is the flow characteristic length.

The current module is in the middle of the first and third modules as it depends in its calculations on the results obtained continuously from the two other modules. The loading information and relative velocities are extracted from the first module. On the other hand the lubricant rheology is continuously corrected using both the loading information and speed information from the first module and the temperature information calculated by the third module. The calculation of temperature distributions in bearing components will be presented in the third module and the temperature at the contact area points will be used to solve the rheology model and to re-calculate the film thicknesses and the resulting shear stresses in the hydrodynamic module.

4.1.3 Bearing temperature calculation – Module 3

For the real surface of a rolling element bearing, the total friction stress is the combination of the results calculated from the two previous modules, i. e. the friction due to solid to solid contact or asperities contact and the friction due to lubricant hydrodynamic effects, shear and drag. The total friction stresses in the bearing can be calculated using any of the real surface micro-contact models such as the Greenwood Williamson, GW model (Greenwood, et al., 1966). However, for the case of the steady and normal operation of a lubricated bearing under constant load and constant speed, it was found that the percentage of asperity contact is less than 5% of the total area of contact. Hence in the normal bearing model, this percentage will be used in Equation (3.51) to calculate the total friction force due to sliding at the contact area defined at any azimuth angle.

The total friction force is used to calculate the heat generation rate due to sliding in the contact area using Equations (3.73) and (3.74), and the drag fiction due to lubricant viscous effect is calculated based on Equation (3.71). The total friction forces and the bearing relative speeds are used to calculate the total heat generation rate at any time step from the beginning of the calculation.

Considering that the system is initially at ambient temperature and the bearing total friction is considered as a fixed source of heat that is uniformly distributed at the circle connecting the centers of the contact areas. Then a thermal network can be constructed for bearing internal parts temperature calculation based on heat transfer models.

4.1.3.1 Theoretical model, lumped thermal capacitance

Solving the transient heat transfer problem to find the temperature of the bearing components and surfaces is a quite complex problem. There is no doubt that the steady state problem is much easier. However, several assumptions can be added to the transient problem to simplify it. One of the major assumptions that can help in solving the problem is to consider the small solid bodies that build the bearing to have uniform temperature distribution at any time. This assumption is known as the lumped mass or lumped capacitance analysis (Lienhard IV, et al., 2003). A dimensionless number, Biot number, is a measure of the lumped mass assumption applicability.. This number specifies whether the heat flow inside the body should be taken into consideration or can be neglected. When this number is very small, < 0.1, the thermal conduction inside body is much higher than the heat flow to outside the body, then the lumped mass analysis can be used (Inceropera, et al., 2007).

The value of this dimensionless number depends on the characteristic length of the body and the ratio between the thermal coefficients of convection and conduction of the body. When the body volume is relatively small and has a high thermal conductivity, the Biot number is very small as shown in Equation (4.6) below

$$Bi = \frac{hL_C}{k} \tag{4.6}$$

where Bi is the Biot number, h is the heat convection coefficient, L_C is the body characteristic length that is defined as the ratio of the body volume and its surface area, and k is the thermal conductivity of the body material.

For the system consisting of the ball bearing, the shaft and the housing, a thermal resistance network will be constructed (Figure 4.2) for transient temperature calculation. Assuming that the system shown in Figure 4.2 is symmetric about the axis of the bearing, only the upper right quarter of the system will be studied.

The system shown in Figure 4.2 can be simplified to 15 thermal nodes with nodes 1 and 6 used for the inner race, nodes 4 and 5 for the outer race, nodes 7 to 10 for the housing, nodes 11 to 15 for the shaft, node 2 for the lubricant and node 3 for the rolling element. The thermal resistance between and two nodes will be named after the numbers of the two nodes, e. g. $R_{1\text{-}6}$ is the thermal resistance between nodes 1 and 6.

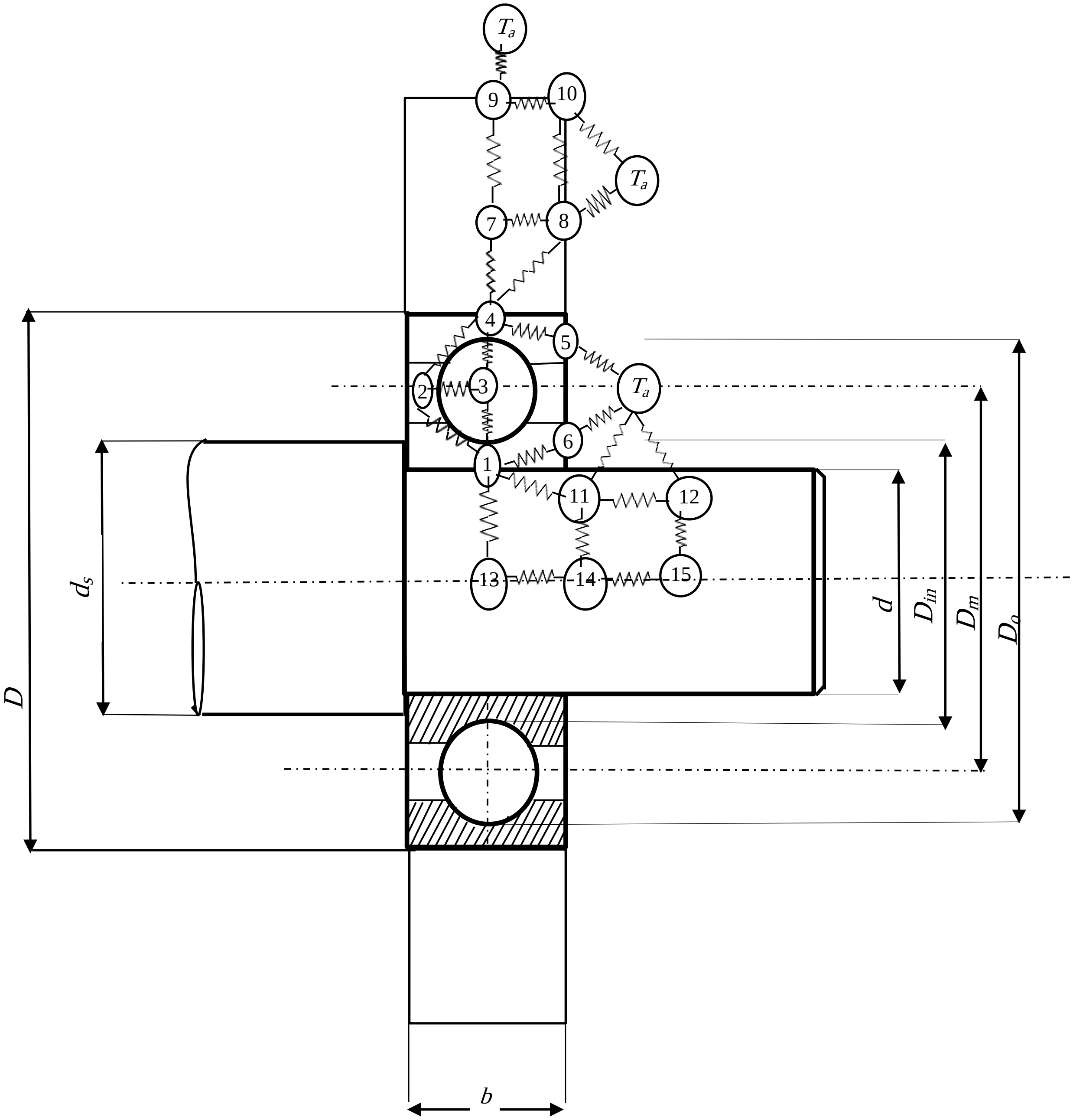


Figure 4.2 The heat transfer network for the radial ball bearing, its housing and the shaft

Also, the ambient temperature for the surrounding is defined as T_a. The heat transfer modes in Figure 4.2 are described in Table 4.1 in which C stands for the conduction heat transfer mode which is the main heat transfer mode between contacting surfaces. The convection and the radiation heat transfer modes are referred to by the letters V and R. The combination of conduction and convection heat transfer modes takes place through the interface area between the rolling element, the inner and outer races and the lubricant grease. All the external nodes, nodes 5, 6, 8, 9, 10, 11 and 12, will be considered to have a combined convection and radiation heat transfer modes. The heat source node, referred to by letter G, is node 3.

Due to the symmetry of the bearing system, shown in Figure 4.2, only one quarter of the system will be modeled. Hence, the value of the heat source used in current module is the quarter of the total heat generation rate calculated by the first and second modules.

Node	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	a
1		C, V	C			C					C		C			
2	C, V		C, V	C, V												
3	C	C, V	G	C												
4		C, V	C		C		C	C								
5				C												V,R
6	C															V,R
7				C				C	C							
8				C			C			C						V,R
9							C			C						V,R
10								C	C							V,R
11												C		C		V,R
12											C				C	V,R
13	C													C		
14											C		C		C	
15												C		C		
a					V,R	V,R		V,R	V,R	V,R	V,R	V,R				

Table 4.1 Heat transfer modes among the thermal network nodes shown in Figure 4.2 where C stands for conduction, V for convection and R for radiation heat transfer modes. The symbol G stands for the heat generation sources.

The heat balance equation for transient temperature of any node in the thermal network can be written as follows:

$$H_{gn} + H_{in} - H_{out} = \boldsymbol{c}\frac{\partial T_n}{\partial t} \tag{4.7}$$

In Equation (4.7), c is the latent heat or specific heat coefficient, the temperature of any node, T_n can be calculated at any time t by calculating the heat generation rate H_{gn} at this node, the heat transferred entering this node H_{in} and the heat dissipated or transferred out from this node, H_{out}.

4.1.3.2 Equations of heat transfer modes and thermal resistances

The heat transfer equations for different heat transfer modes are defined as follows (Inceropera, et al., 2007):

1. Conduction mode:

$$H_{cond} = \frac{\boldsymbol{a}}{\boldsymbol{\ell}}\boldsymbol{k}\left(T_1 - T_2\right) \tag{4.8}$$

2. Convection mode:

$$H_{conv} = h_v \mathcal{S}\left(T_1 - T_2\right) \tag{4.9}$$

3. Radiation mode:

$$H_{Rad} = 5.73\varepsilon\mathcal{S}\left[\left(\frac{T}{100}\right)^4 - \left(\frac{T_a}{100}\right)^4\right] \tag{4.10}$$

In Equation (4.8) the conduction heat flow rate H_{cond} between two points that have temperatures T_1 and T_2 is presented, where $\boldsymbol{k}$ is the thermal conductivity (watt/m.K), $\boldsymbol{a}$ is the area through which the heat flow is passing and $\boldsymbol{\ell}$ is the distance between the two flow points. In Equation (4.9) the heat convection rate H_{conv} depends on, in addition to the temperature difference, the surface area from which the heat flows, $\mathcal{S}$, and the coefficient of heat convection, h_v.

In the case of rolling element bearings, the calculation of convection heat coefficient is very complex especially for the bearing internal parts in the presence of a lubricating fluid. Therefore, the convection heat coefficient is calculated using some empirical formulas. Each one of these formulas is found to be suitable for certain heat flow conditions. For the case of the heat convection inside a lubricated rolling element bearing assuming forced convection and laminar flow, the Pohlhausen solution leads to the Eckert formula for calculating the heat convection coefficient (Pitts, et al., 1997). This formula gives the heat convection coefficient h_v based on the Reynolds number, R_e, of the flow between the ball and the bearing race, the ball diameter D_b,,the fluid thermal conductivity $\boldsymbol{k}$ and the Prandtl number, Pr (Harris, et al., 2007), i.e.,

$$h_v = 0.0332 \frac{\boldsymbol{k}}{D_b} \ \mathrm{Pr}^{0.4} \sqrt{\mathrm{Re}} \tag{4.11}$$

The free convection average coefficient between bearing, shaft and housing external surfaces and the ambient surrounding is given by the following equation:

$$\bar{h}_v = \frac{\boldsymbol{k}}{D} \mathrm{Re}^{m} \ \mathrm{Pr}^{1/3} \tag{4.12}$$

where Pr is the Prandtl number , *Re* is the Reynolds number and *m* is an empirical factor that can be found in (Pitts, et al., 1997) for different types of geometries.

For the heat radiation equation, i.e., Equation (4.10), the rate of radiated heat from a surface area S at absolute temperature T to a surrounding with an ambient absolute temperature T_a, the radiating surface has an emissivity, ε. The emissivity coefficient value for the materials of bearing, shaft and housing are listed in (Inceropera, et al., 2007).

Thermal resistance for the system shown in Figure 4.2 can be divided into the following three groups:

1. Thermal resistance for radial conduction between shaft points $R_{1\text{-}13}$, $R_{11\text{-}14}$ and $R_{12\text{-}15}$ $\mathrm{R}_{3\text{-}4}$, $\mathrm{R}_{3\text{-}1}$, e.g.

$$R_{1-13} = \frac{\ln\left(\dfrac{r_{13}}{r_1}\right)}{2\pi \mathscr{k} l} \tag{4.13}$$

2. Thermal resistance for linear conduction $R_{1\text{-}6}$, $R_{1\text{-}11}$, $R_{11\text{-}12}$, $R_{13\text{-}14}$, $R_{14\text{-}15}$, $R_{4\text{-}5}$, $R_{4\text{-}7}$, $R_{4\text{-}8}$, $R_{7\text{-}8}$, $R_{7\text{-}9}$, $R_{8\text{-}10}$, $R_{9\text{-}10}$ and $R_{1\text{-}6}$, e.g.

$$R_{1-6} = \frac{\Delta l}{\mathscr{S}\mathscr{k}} \tag{4.14}$$

3. Thermal resistance for convection heat transfer $R_{1\text{-}2}$, $R_{3\text{-}2}$, $R_{4\text{-}2}$, $R_{5\text{-}a}$, $R_{6\text{-}a}$, $R_{8\text{-}a}$, $R_{9\text{-}a}$, $R_{10\text{-}a}$, $R_{11\text{-}a}$, and $R_{12\text{-}a}$:

$$R_{conv} = \begin{cases} \dfrac{1}{(h_v)_{air}\,\mathscr{S}} & R_{5\text{-a}},\ R_{6\text{-a}},\ R_{8\text{-a}},\ R_{9\text{-a}},\ R_{10\text{-a}},\ R_{11\text{-a}},\ \text{and } R_{12\text{-a}}. \\ \dfrac{1}{(h_v)_{oil}\,\mathscr{S}} & R_{1\text{-}2},\ R_{3\text{-}2} \text{ and } R_{4\text{-}2} \end{cases} \tag{4.15}$$

4.1.3.3 Computational algorithm

The equations and formulas defined in the above section are used to calculate the heat flow rates and temperature variations for each lumped mass defined as a node on the thermal network as shown in Figure 4.2. The temperature calculated at each thermal node will be used for re-calculation for all the processes that are affected by temperature variations in all modules. For example, the rate of heat generation will be recalculated at each time step to compensate for the change in the lubricant viscosity which affects both the sliding friction in the lubricant fluid film and the drag friction. Also, H_{in} and H_{out} will be calculated at each time step using heat transfer equations and resistance definitions as shown in the previous section. The heat balance equations for each node will be iteratively calculated until the lubricant viscosity difference due to temperature change is smaller than a specified limit. At each time step the temperature of each node will be calculated and then used to solve the heat balance equation for the next time step. The algorithm that is used for the third module can be summarized by the following steps and is depicted by the flowchart in Figure 4.3

1. Calculate bearing heat generation rate in transient status before reaching thermal equilibrium condition, considering steady speed and constant load.
2. Calculate heat dissipations from different bearing surfaces.
3. Calculate surface temperature for a specified surface of the bearing resulting from the net heat balance, after subtracting the dissipated heat from the generated heat.
4. Recalculate the heat generation and the heat dissipation and the net heat balance using the modified lubricant viscosity due to temperature change.
5. Repeat the previous step until the difference in the lubricant viscosity due to the temperature change is less than a pre-specified small value.
6. Draw the temperature rise curve for a specified bearing surface from the initial condition to the thermal equilibrium condition.

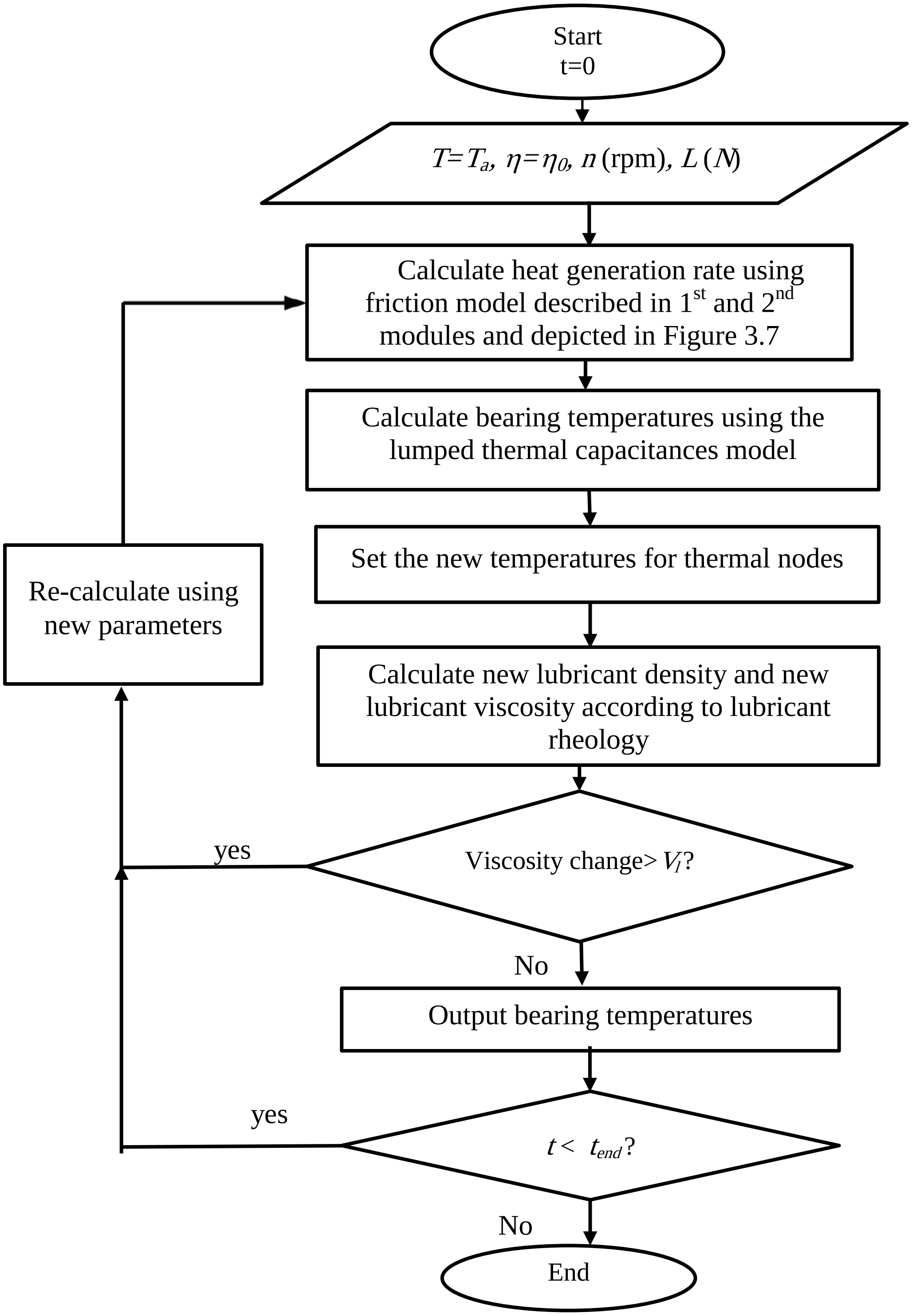


Figure 4.3 Flowcharts for bearing temperature calculation

4.2 Heat Generation Model Verification

In this section, two famous bearing friction calculation models, i.e., Palmgren (Palmgren, 1959) and SKF (SKF, 2003) models, will be used to validate the results obtained by the resulted sum of the first and second modules of the proposed model. The comparison will be accomplished for bearing steady state operation at different loading and speed conditions. The used bearing parameters will be the same for the proposed model and the two compared models. A list of the test bearing parameters is shown in Table 4.2 below.

The friction and heat generation rate for the SKF 6202-16RSH bearing, will be calculated for the proposed model and the two empirical models at five different radial loads: 10, 20, 30, 40 and 50 Newton. For each loading condition, the friction and heat generation rates will be calculated at different steady speeds ranging from 1800 rpm to 12000 rpm. The lubricant used in the test bearing is LUMAX-1911 grease whose specifications are shown in Appendix A. The lubricant viscosity at zero load and at ambient temperature is ν_0 = 40 cSt

Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Bore diameter, d_{sh}	16 mm	Inner race diameter, D_{in}	20.728 mm	Ball diameter, D_b	6 mm
Outer diameter, D	35 mm	Mean (pitch) diameter, d_m	26.734 mm	Number of balls, Z	8
Width, b_w	11 mm	Outer race Diameter, D_o	32.74 mm	Radial clearance, P_d	12 µm

Table 4.2 Parameters of the test bearing (Type:, SKF 6202-16).

The two models are originally established to calculate bearing friction torques and heat generation rates at thermal equilibrium condition. Hence, the first validation of modules 1 and 2 of the model will be carried out in a steady state environment. This means that the calculation will start at a certain assumed temperature distribution that will be used in the model to determine the lubricant properties during the steady state operation. Also, the friction shear stress will have a steady distribution which depends only on the loads and speeds of the points in the contact area. That means after the bearing reaches a steady state temperature. However, in this section the two empirical methods will be used to calculate friction and heat generation during temperature rise of bearing.

The use of the Palmgren model to find transient temperature rise rather the temperature of equilibrium condition was tested and its output was compared with experimental results previously in some research works such as (Jeng, et al., 2003; Takabi, et al., 2013). The lubricant viscosity will be updated for every time step and then used to re-calculate the friction torque and the heat generation rate. The calculation will be iterated for each time step until the calculated viscosity converges. The results of proposed model in sections 4.1.1 and 4.1.2 will be compared with the results of the two modified empirical models when applying three different radial loads and then when running the bearing at different steady speeds ranging from 1800 to 12000 rpm.

4.2.1 Verification by the Palmgren model

This model was described in Chapter 3 by Equations (3.3) to (3.10). According to Equation (3.4), the value of the product $v_0 \times n$ will specify the equation to be used to calculate viscous friction torque M_v.

Speed, rpm	$v_0 \times n$	Viscous drag friction torque, $Mv_{,,}$N.mm	Speed, rpm	$v_0 \times n$	Viscous drag friction torque, Mv, N.mm
1800	72×10^3	4.95	7200	288×10^3	12.63
3000	120×10^3	7.05	9000	360×10^3	14.67
4800	192×10^3	9.67	10800	432×10^3	16.56
6000	240×10^3	11.19	12000	480×10^3	17.745

Table 4.3 Viscous friction torque for different speeds

Table 4.3 shows the resultant viscous friction torque when considering different speeds. Each calculation is done for the steady speed used for each calculation step.

The load dependent friction torque, M_l, will be calculated by

$$M_l = f_1 F_\beta d_m \tag{4.16}$$

In the current case study, the load is a pure radial load which can be 10, 20 or 30 Newton and the bearing coefficient f_1 is given by Equation (3.8). When z=0.0005, y=0.55, as given in (Palmgren, 1959) for radial deep groove ball bearing. The equivalent load $F_s = F_r$ and the static load rating, C_s= 3750 N, as given in the SKF bearing catalogue (SKF, 2003). Then the friction torques for different loading conditions after considering the effect of coupling misalignment are:

Load	10 N	20 N	30 N	40 N	50 N
Friction torque N.mm	0.066	0.133	0.20	0.26	0.33

Table 4.4 Friction torque for different loading cases obtained by the Palmgren Model

The total friction torque is the sum of the results shown in Tables 4.3 and 4.4. The total heat generation rate is calculated using the Palmgren model (Equation (3.10)) and its values for different load and speed combinations are shown in Table 4.5.

Speed \ Load	10 N	20 N	30 N	40 N	50 N
1800	0.957	0.97	0.98	0.99	1.00
3600	3.025	3.051	3.076	3.10	3.12
5400	5.935	5.97	6.01	6.04	6.08
7200	9.576	9.62	9.677	9.727	9.77
9000	13.88	13.94	14.00	14.06	14.1
10800	18.79	18.874	18.950	19.025	19.101
12600	24.29	24.38	24.47	24.56	24.64

Table 4.5 Heat generation rates obtained by the Palmgren model (in watts)

4.2.2 Verification by the SKF model

As the test bearing is grease lubricated, only three types of friction moments will be calculated using Equations (3.12) to (3.14). These three constituents of the total friction moment are calculated as follows:

4.2.2.1 Rolling friction moment M_r

The coefficient of rolling friction G_{rr} for radially loaded deep groove ball bearing is calculated using the following formula (Croes, et al., 2009):

$$G_{rr} = R_1 d_m^{1.96} F_r^{0.545} \tag{4.18}$$

where the value of R_1 is found in (SKF, 2003) for the deep grove ball bearing series 62XX and equals 3.9×10^{-7} . The rolling friction moment at certain thermally steady state can be calculated by the following equation:

$$M_r = G_{rr} \left(\nu \cdot n\right)^{0.6} \tag{4.19}$$

where ν is the kinematic viscosity of the lubricant at the operating temperature and n is the rotational speed in revolutions per minute. Table 4.6 shows the values for rolling friction moment considering the given loading and speed conditions.

Load / Speed	10 N	20 N	30 N	40 N	50 N
1800	0.468	0.603	0.7164	0.816	0.906
3600	0.637	0.819	0.973	1.237	1.37
5400	0.844	1.086	1.29	1.578	1.75
7200	0.965	1.24	1.47	1.875	2.083
9000	1.07	1.38	1.64	2.143	2.38
10800	1.2	1.58	1.88	2.39	2.65
12600	1.37	1.76	2.099	2.62	2.91

Table 4.6 Rolling friction moments (in N.mm) obtained using the SKF model for different loading and speed conditions.

4.2.2.2 Sliding friction

Referring to Equation (3.13) and considering the constants for the test bearing from the bearing catalogue (SKF, 2003), μ_{sl} equals 0.04 and the sliding friction coefficient, G_{sl} can be calculated at different loading conditions as follows:

$$G_{sl} = S_1 d_m^{-0.26} F_r^{5/3} \tag{4.20}$$

where S_1 is a geometry constant for bearing sliding friction. For deep groove ball bearing series 62XX, S_1 equals 3.23×10^{-3} (SKF, 2003). In the case of mixed lubrication when there is a possibility of dry friction, the sliding coefficient of friction, μ_{sl} can be calculated using the following equation (SKF, 2003):

$$\mu_{sl} = \varphi_{bl}\mu_{bl} + \left(1-\varphi_{bl}\right)\mu_{EHL} \tag{4.21}$$

where μ_{bl} is a lubricant property factor ≈0.15, μ_{EHL} is the friction coefficient in full film condition and φ_{bl} is a weighting factor obtained based on the lubricant viscosity and bearing dimensions given below (SKF, 2003):

$$\varphi_{bl} = \exp\left(-2.6\times10^{-8}\left(n\nu\right)^{1.4} d_m\right) \tag{4.22}$$

where n is the rotational speed in rpm, ν is the lubricant kinematic viscosity in cSt and d_m is the mean diameter.

The values for sliding friction moments (in dyne.mm) at different load and speed conditions are shown in Table 4.7. It is clear that this friction component does not depend on rolling speed. The total heat generation rates calculated by the SKF model for rolling and sliding frictions are listed in Table 4.8.

Load Speed	10 N	20 N	30 N	40 N	50 N
Any	622	1345	2275	3389	4675

Table 4.7 Sliding moments (in dyne.mm) obtained using the SKF model for different loading and speed conditions.

Load Speeds	10 N	20 N	30 N	40 N	50 N
1800	4.634	4.661	4.684	4.705	4.724
3600	9.36	9.43	9.50	9.56	9.62
5400	14.15	14.302	14.43	14.54	14.65
7200	18.99	19.23	19.43	19.619	19.78
9000	23.89	24.23	24.519	24.77	25.01
10800	28.83	29.284	29.66	30.012	30.32
12600	33.81	34.39	34.88	35.32	35.72

Table 4.8 Heat generation rates (in watts) obtained using the SKF model for different loading and speed conditions.

4.2.2.3 Seal friction

Seal friction is also considered in the SKF model. This is can be considered as an advantage over the Palmgren model which does not count for the seal friction. In the SKF model the heat generated by seal friction can be calculated using Equation (3.14). This equation has three constants, K_{s1}, K_{s2} and γ, were previously determined by curve fitting of the experimental work. According to the test bearing type and manufacturer data, the values for these constants are given in Table 4.9 for the RSH seal type and deep groove SKF 6202 ball bearing.

K_{s1}	0.028	K_{s2}	2
γ	2.25	d_s=d_2 (seal counter face diameter)	19.4 mm

Table 4.9 Constants for calculating seal friction moment for single row deep groove ball bearing (Type: SKF 6202-2RSH)

Using the values shown in Table 4.9, the values for seal friction of this type of bearing can be estimated as a constant value that should be added to the other friction moments. For the test bearing this value is calculated and it equals 24.12 N.mm for the two seals at both bearing sides.

4.2.3 Comparison of the results obtained by the Palmgren, SKF and the proposed friction models

In the current section, first and second modules of the proposed friction model will be used to calculate the heat generation rate for the same bearing under the same operating conditions shown in the previous two sections. The current comparison will be accomplished at thermal steady state condition. The comparison of the friction moments calculated by different methods is not possible because each of the methods has different definition for each friction component. This difference in frictions definition resulted in the difference of the factors affecting each type of friction, calculated by each method. A meaningful comparison can be done by examining the total friction or the total heat generation rates resulted from the three methods under different load and speed conditions. As the SKF model is the only model that calculates the seal friction, the comparison will be for the heat generation in the rolling element/raceway contact. For accurate calculation of the bearing temperatures, the value of the seal friction yielded by the SKF model will be added to the total friction calculated by the two other models.

4.2.3.1 Comparison at the thermal equilibrium condition

In this comparison, the heat generation rate is calculated considering that the bearing has reached its thermal equilibrium state, i.e., the heat generated due to bearing operation equals the heat dissipated to the surrounding. In this state there is no change in the lubricant properties due to the rise of the temperature.

The comparison will be accomplished at two different loading conditions at speeds ranging from 1800 rpm to 12600 rpm. The two loading conditions are the light loading (10, 20, 30, 40 and 50 Newton) and the moderate loading (50, 150, 250, 350 and 450 Newton). Figures 4.4 and

4.5 present the total heat generation rates obtained by the three models for the two loading conditions, respectively.

The following can be observed from the two figures:

1. For the light loading conditions, the proposed model has values of heat generation rate between those of the other two models.
2. The proposed model is more sensitive to the load condition changes than the other two models. The Palmgren model is the least sensitive to load changes. This may suggest that the proposed model is more accurate because the load change should cause changes in heat generation rate.
3. Under the moderate loading conditions, all three models exhibit sensitivity but the difference of the generated heat at different loads is much more recognizable in the results of the proposed model.
4. All the three models indicate that the heat generation rate increases with the increase of the running speed for any loading condition.

The effect of the speed on the heat generation rate of the three friction components is shown in Figure 4.6. It can be seen from this figure: a) all the heat generated from all the three sources increase with speed, b) the contribution of the drag friction increases faster than those of the other two sources, and c) the contribution of the drag friction surpasses that of dry friction.

The load effect on the heat generation contributions of the three sources is shown in Figure 4.7. Contrary to the speed effect, the load increase when running at the same speed has virtually no effect on the heat generated by drag friction in this loading range. The load change effects on the heat generation behaviours of the dry and hydrodynamic shear frictions are similar to but at lower rate than those observed in speed change conditions.

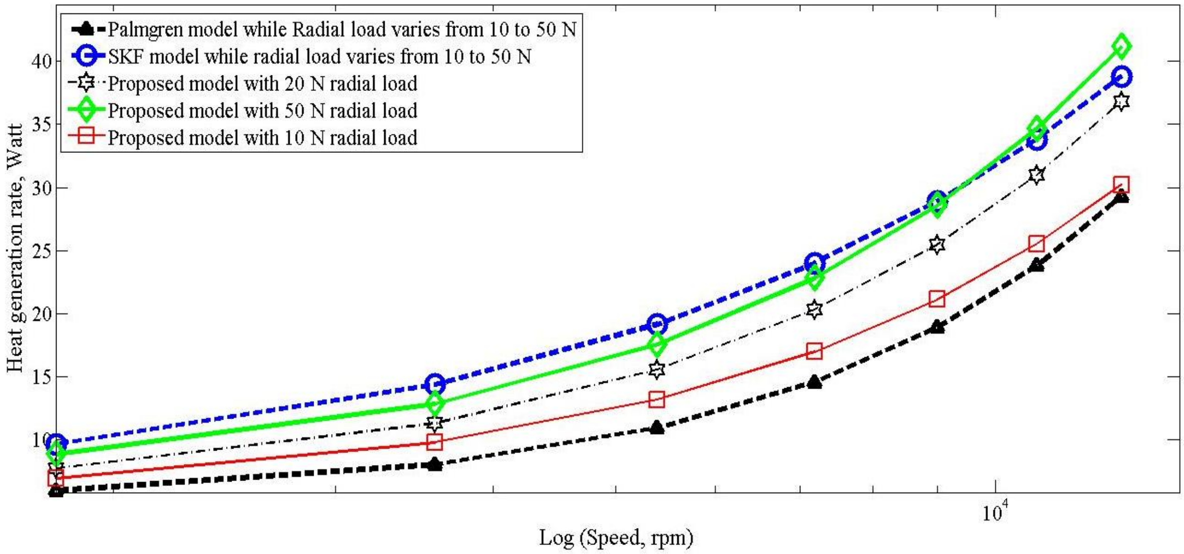


Figure 4.4 Heat generation rates obtained by the three different models: the Palmgren model, the SKF model and the proposed model, calculated at low loading condition, 10 to 50 N radial loads.

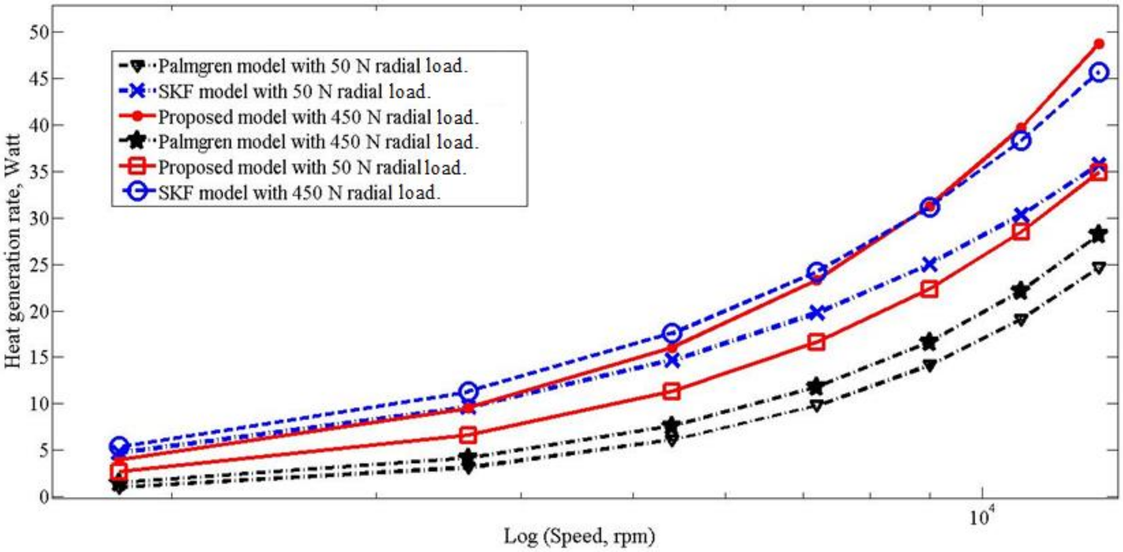


Figure 4.5 Heat generation rates obtained by the three different models: the Palmgren model, the SKF model and the proposed model, calculated at moderate loading condition, 50 to 450 N radial loads.

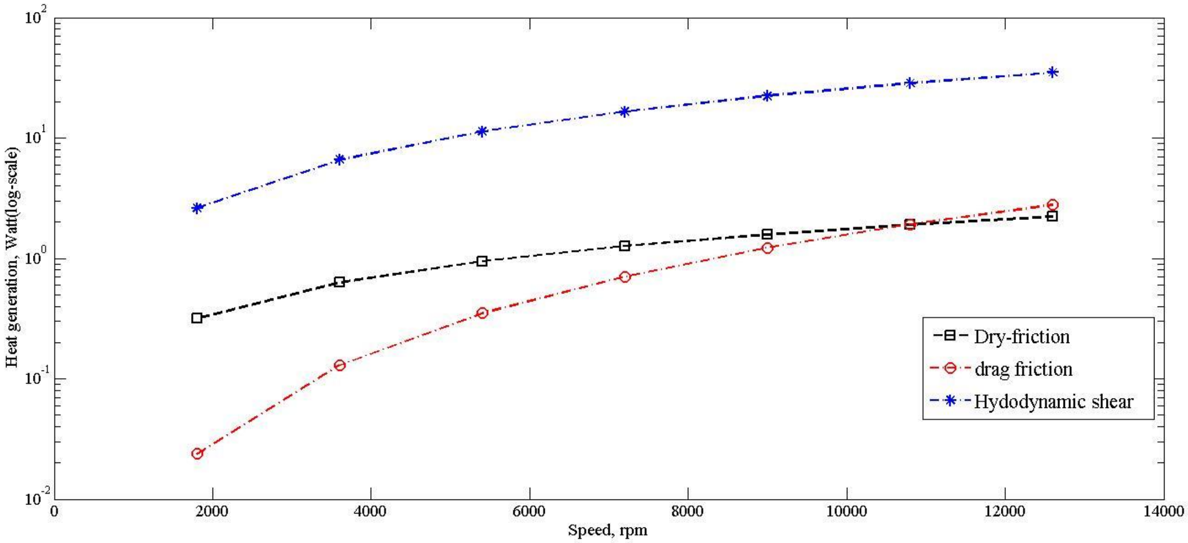


Figure 4.6 Effect of speed on different heat generation sources as defined by the proposed model

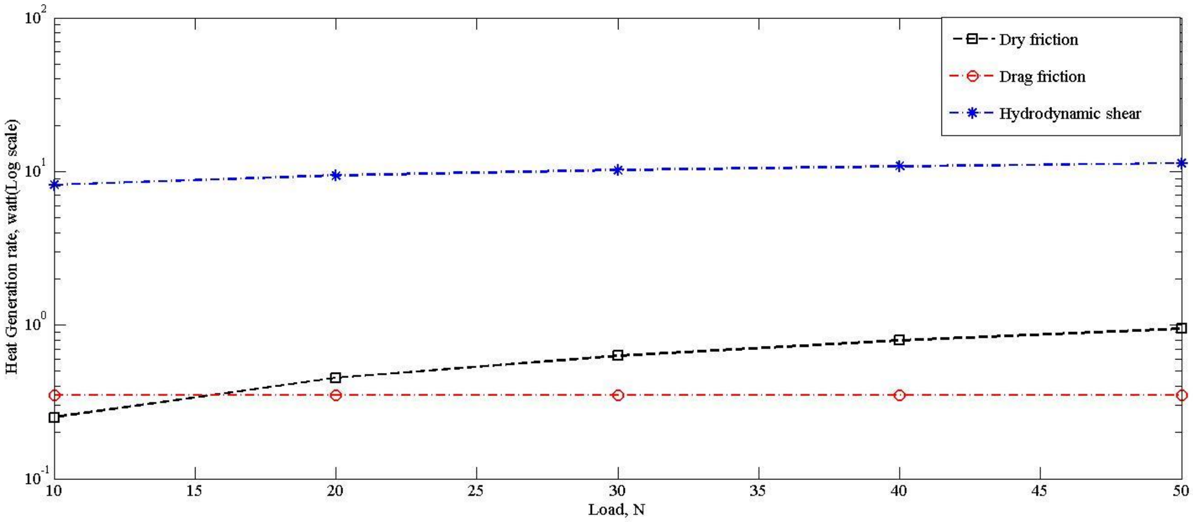


Figure 4.7 Effect of the load on the different heat generation sources as defined by the proposed model.

4.2.3.2 Comparison at the thermal transient condition

Under the bearing thermal transient condition, the lubricant properties change continuously causing a change of both the heat generation and heat dissipation rates. This change in heat balance equation at every moment during the transient period is different for every tribology system and is very difficult to quantify. Although the bearing friction models were originally established for calculation of equilibrium temperatures, they were used in some references to calculate the temperature rise curve during the transient period (Jeng, et al., 2003; SKF, 2003). The effect of lubricant rheology on the calculated bearing temperatures can be found in the three models as follows:

1. Palmgren Model: as described in section 4.2.1 there are two friction sources in this model: the load dependent and the viscous dependent friction sources. Only the second friction source can change with lubricant viscosity changes at different temperatures if the ambient viscosity ν_0 in Equation (3.4) is replaced with a temperature dependent viscosity.
2. SKF Model: as described in section 4.2.2 there are three friction sources in this model. The rolling friction (Equation (4.19)) and the mixed lubrication sliding friction (Equation (4.22)) are affected by the lubricant viscosity.
3. The proposed analytical model: the lubricant viscosity is a main factor that contributes in the following calculations:
    a) Hydrodynamic shear friction, Equations (3.61), (3.50 a) and (3.50 b).
    b) Viscous drag as it will affect the Reynolds number, Equations (4.4) and (4.5).
    c) Mixed lubrication ratio of dry friction, equations for the GW model.
    d) Heat dissipation through heat convection coefficient, Equation (4.11)

4.3 Bearing Temperatures, Heat Transfer Model Verification

To validate the third module of the proposed heat transfer model, the heat generation rate calculated by first and second modules of the model will be used as an input to the third module to calculate bearing temperatures during the transient state. Then the calculated temperatures will be used to re-evaluate the heat generation rate as was described in the computational algorithm flowchart (Figure 4.3).

4.3.1 Test equipment

The calculated temperatures will be compared to the measurements acquired experimentally using thermocouple (for outer race and housing), infrared sensor and thermal camera (for rotating parts, the inner race and the shaft). The experiments are conducted on a high speed bearing test rig that has both temperature and vibration measuring devices. The test rig is driven by a high speed motor having a speed range up to 20,000 rpm. The speed of the motor is controlled using a SJ-200 Hitachi VFD (variable frequency drive). The load effect will be studied by the measurements from three radial loading steps. Each load step is a 1 kg disc that is clamped to the shaft using a shaft-hub locking device. This device is a good tool of clamping loads at any location without using a keyway or thread holes in the shaft. The shaft has two steps, 16 mm under the bearings and 22.225 mm for the loading portion. A HD-810 Magtroll dynamometer is connected to the shaft to apply and control torque load. The interfaces between the motor and the shaft, and between the dynamometer and the shaft are achieved using high speed anti-vibration flexible couplings. A FLIR E-40 thermal camera is used for thermal video recording. Figure 4.8 a, b and c, shows photos of the test rig and the HD-810 Magtroll dynamometer corresponding to the three loading steps. Two kinds of accelerometers are used for measuring the vibration signal in two angles, the vertical and the horizontal axes. The first accelerometer is installed at vertical angle and it is a high frequency accelerometer (Type ICP-352A60) that can detect frequencies up to 60 KHz. The other accelerometer is installed at 90° or horizontal angle and it has a detectable frequency limit of 20 KHz. Before entering the data acquisition system, the output of the accelerometers goes first through a signal conditioner device that is a kind of low pass filter that also acts as anti-aliasing filter.

In Figure 4.9a schematic drawing for the bearing test rig is presented. The thermal camera is set to record the temperature changes on the bearing side surface by choosing the Rectangle Box option of the thermal camera. This option allows recording maximum and minimum temperatures in the chosen area in addition to the ability of measuring the temperature at one point using the Spot option. In order to suppress the high reflection effect of the unwanted areas, the bearing housing and the shaft areas are covered by low reflectivity tape.

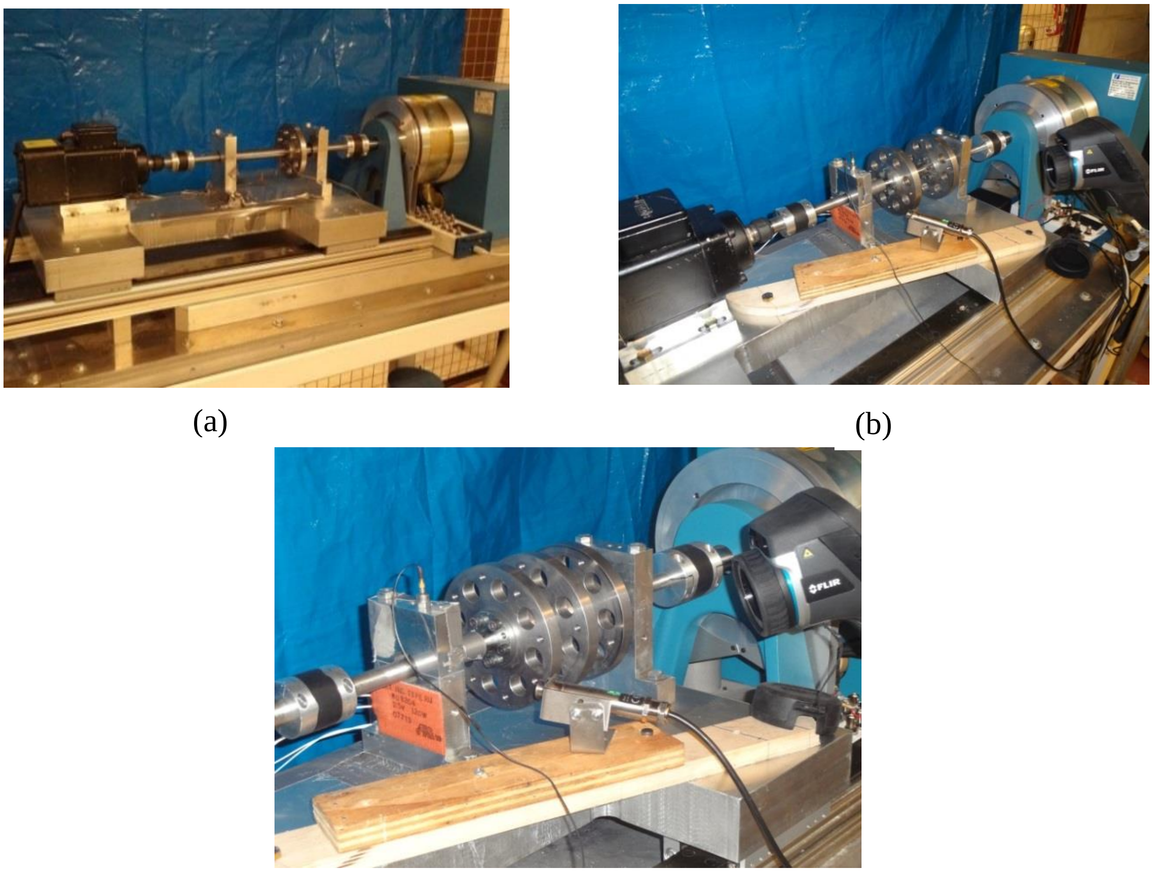
4.3.2 Surface temperature measurement devices

In this study, two non-contact temperature measurement devices are used to measure the temperature of the bearing surface. These devices are the E40 thermal camera supplied by the FLIR company and the OS211 infrared sensor produced by OMEGA incorporation. Although the two devices are designed to measure the same type of energy, radiation energy, and they almost work in the same spectral range, some other factors should be taken into consideration when their readings need to be compared. Using these two different devices is essential to get more confident readings of bearing surface temperatures. The thermal camera is more flexible than the infrared sensor as the emissivity coefficient for the targeted surface in addition to the target distance can be adjusted and varies from 0.01 to 1 for the emissivity. The infrared sensor does not have this capability. The emissivity for the surface whose temperature is measured by this sensor should be close to the specified value of 0.95. This limitation for surface emissivity makes the sensor suitable only for measuring temperature of limited kinds of surfaces such as rubber, food, leather, plastic and textile.

In order to measure surface temperature for metallic or polished surfaces, an external function should be added to correct the measurement due to emissivity difference. The advantage of the infrared sensor is its smaller volume and lighter weight than the thermal camera. This advantage can make the insertion or the embedding of the sensor into machine cases more easily in the future. Another advantage of the infrared sensor is its convenient output as it has the same output as that of a K-type thermocouple. This advantage makes it very easy to be used with several commercial types of data acquisition systems and software. The outputs of both the thermocouple and the infrared sensor are acquired and converted from voltage to temperature readings using the data acquisition system and the LabView software. The Labview software contains several built-in functions that accept the data from thermocouples of different types. The main characteristics of the two temperature monitoring devices are detailed in Table 4.10 (Flir Incorporation, 2011; Omega Incorporation, 2013).

Parameter	**Thermal Camera FLIR E-40**	**Infrared sensor OMEGA OS211**
Minimum focus distance	0.4 m	0.1 m
Thermal sensitivity	< 0.07 ºC at 30 ºC	0.5 °C
Image frequency	60 Hz	240 ms (4.2 Hz)
Temperature range	0 ºC to 650 ºC	-20 ºC to 500 ºC
Accuracy	±2 °C or ± 2% of reading	±1 °C or ± 1% of reading
Emissivity	Adjustable from 0.01 to 1	95% (fixed)
Spectral range	7.5 to 13 µm	8 to 14 µm
Output	Thermal image and thermal videos	Thermocouple output
Weight	825 g	90 g

Table 4.10 Main features of the E-40 thermal camera and OS211 infrared sensor



(a) (b)

(c)

Figure 4.8 Bearing temperature test rig with different loading setups: (a) single load disc, (b) two load discs, and (c) three load discs.

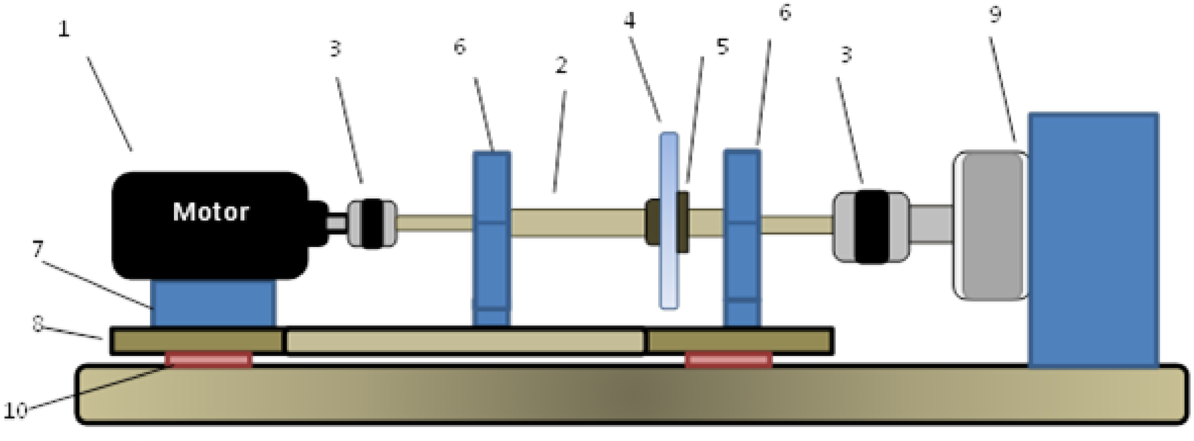


Figure 4.9 1-AC motor (up to 20,000 rpm), 2-Two-step 1566 steel shaft, 3-Anti-vibration flexible couplings (2), 4-Radial load (Disc), 5-Shaft-Hub Locking Device (SHLD).6-Bearing and bearing Housings (2 each), 7-Motor base, 8-Sliding base, 9-HD-810 dynamometer and 10-Sliders (4), 11-FLIR E40 thermal camera (Not shown), 12-Infrared sensor (Not shown) and 13- thermocouples at bearing outer race (2) (Not shown) .

4.3.3 Normal bearing tests

The accurate bearing transient temperature curves when at normal condition, i.e. no faults exist, are an important result that should be validated before using the same analytic model to calculate transient temperature curves for certain faulty bearing conditions, that will be shown in Chapter 6.

In the current section, the transient temperature rise curves for the bearing inner and outer races external side surfaces are measured by the thermocouple for the outer race, the thermal camera and the infrared sensor for the inner race are recorded. Then the temperature rise curves for the nodes in Figure 4.2 that represent the same surfaces are obtained considering the same loading and speed conditions, through two calculation cases. The first case considers constant rate of heat generation and no change in lubricant viscous properties. In other words, the calculation is accomplished considering a fixed heat source and a constant lubricant rheology while the temporal development of bearing temperature is only due to the changes in heat dissipation rates as a result of the system nodes temperature rise. In the second case of calculation, the numerical algorithm described in Figure 4.3 for transient temperature calculation taking into account the changes of lubricant properties due to temperature rise and the resulting temporal changes in bearing heat generation rates is applied. The results for these two calculation cases for the proposed analytic model, the three modules, and the semi-empirical models (SKF or Palmgren with the third module for heat transfer modeling) in addition to the experimental measurements for both inner and outer races are shown in Figures 4.10 and 4.11.

There are two purposes for which the calculations using the three methods are repeated for the two different cases. The first is to examine the sensitivity of the analytical model to the factors affecting the heat generation and heat dissipation rates while the second is to compare this sensitivity to those of the existing SKF and Palmgren models when used to measure transient temperatures instead of their original function of bearing heat calculation at equilibrium condition.

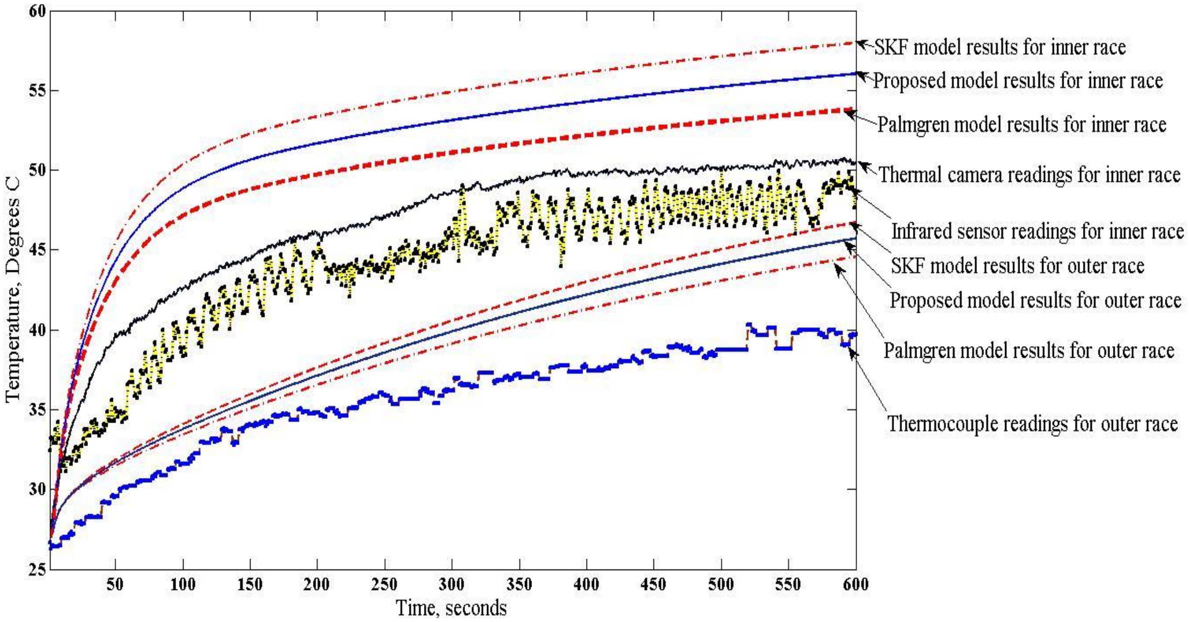


Figure 4.10 Comparison between the temperature rise curves for inner and outer races surfaces as measured using different measuring devices and as calculated using the heat transfer model in accordance with the heat generation rate calculated by SKF model, Palmgren model and the proposed model

The results of the first calculation case in addition to experimental measurement of bearing inner and outer race temperature rise curves are shown in Figure 4.10. It can be seen from the figure that there is a difference between the calculated and the measured temperatures for all of the three temperature calculation methods. This difference is larger for the point with higher temperature, the inner race, than the point with lower temperature, the outer race. This larger difference for the hotter surface can be explained as a result of faster change of lubricant properties at locations that are closer to the hotter surfaces. This fast change causes a big difference between the actual measured temperature and the calculated temperature when no change in lubricant properties is considered. However, the trend of temperature curve for all of the three models is the same. This is because the heat transfer module uses the output of friction generation models as a single value. For accurate calculation of the bearing temperatures, the effect of lubricant temperature rise must be taken into consideration as described in the computational algorithm in Figures 3.7 and 4.3. The effect of temperature variation on lubricant viscosity and thermal properties, thermal conductivity and specific heat in addition to the effect of shear rate of the grease on its effective film thickness are shown in detail in Appendix A. Another finding from Figure 4.10 is the strong fluctuations of thermocouple and infrared sensor readings. This fluctuation is caused by the electromagnetic noise resulted from the revolution of ferromagnetic material which is strong enough to cause this fluctuation in low voltage that represents the measured temperature. As the thermal camera is isolated from this effect, its readings are smoother.

The resulted temperature curves for the second case of calculation are shown in Figure 4.11. In this case of calculations, the variations in lubricant properties are used to calculate the heat generation and the heat dissipation rates for each time step as explained in Section 4.1. Figure 4.11 shows that the proposed analytic model could yield transient temperature curves that are the closest to the measured transient temperature curves. Although the Palmgren model yields a lower heat generation rate than the proposed model, as depicted in Figures 4.4 and 4.5, it can be seen from Figure 4.11 that the temperature rise curves calculated by this model is higher than that calculated by the proposed model after considering heat generation variation. The explanation of this result again indicates the poor dependency of the Palmgren model on lubricant properties as can be concluded from Equations 3.4 and 3.5. On the other side, the heat generation rate calculated by the proposed model is more sensitive to lubricant property changes as it includes in its computation a higher level of dependency on lubricant properties. In other words, for the proposed model the lubricant properties are included in the calculation of lubricant film thickness, hydrodynamic shear stress and lubricant viscous drag which save a more accurate calculation for transient heat generation rate. Figure 4.11 also shows that, using the SKF bearing heat generation model with the heat transfer module produces transient temperature curves that are higher than the measured curves. It was also previously shown that the SKF model led to the highest heat generation rates compared to the other two methods as depicted in Figures 4.4 and 4.5

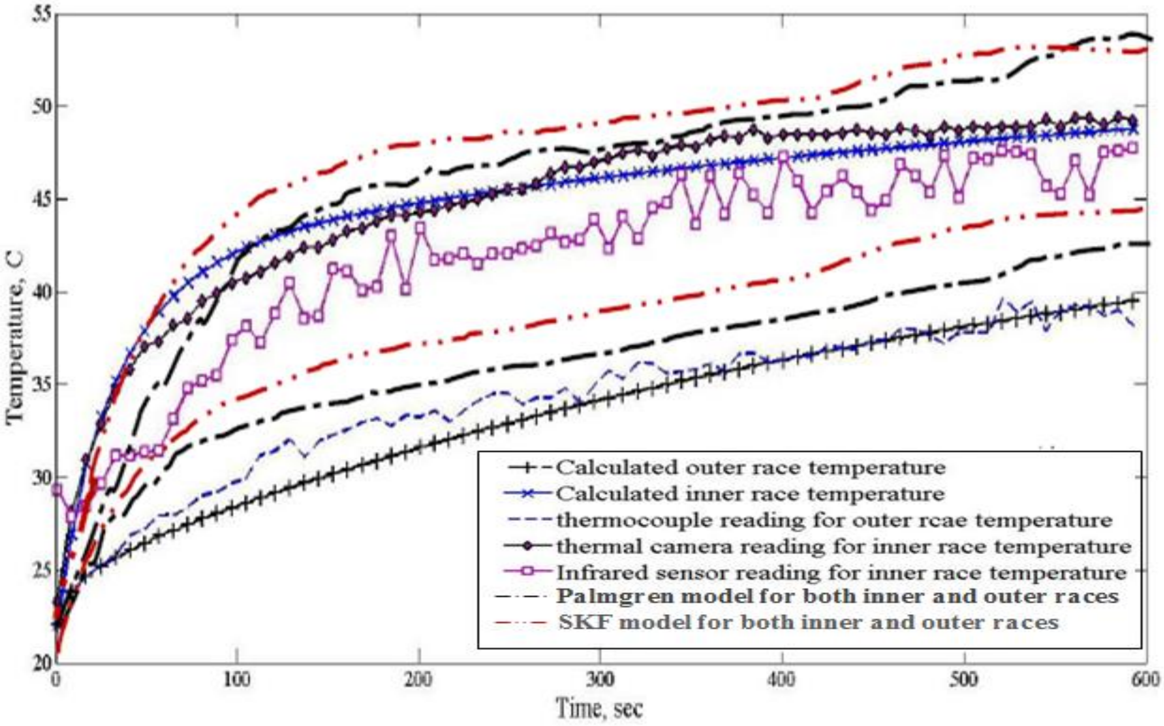


Figure 4.11 Comparison between the temperature rise curves for inner and outer races surfaces as measured using different measuring devices and as calculated using the adaptive bearing temperature model.

Finally, from the above discussion and the results depicted in the two figures, 4.10 and 4.11, it can be concluded that the proposed analytic model can accurately calculate bearing transient temperature curves at least at the load and speed ranges that were tested in previous sections. Then the proposed model is now ready to be used for the calculation of the temperature rise curve after modifying the heat generation modules to account for some bearing faults. The bearing faults effect on the bearing transient temperature model will be discussed and examined for real bearing faults in Chapter 6.

Chapter 5

Vibration-Based Bearing FaultDetection:Envelope Analysis

The envelope analysis method is one of the most widely-used methods for detecting impulsive bearing faults. The effectiveness of this method depends on two factors:

1. The envelope window selection.
2. In-band noise filtering and in-band signal enhancement.

The envelope window selection is carried out through the optimization of a band pass filter parameters, center frequency and band width. There are several methods in the literature discussed the selection of these parameters. In the current chapter the recent-defined statistical property, the spectral kurtosis, SK, will be used for this purpose.

The in-band noise filtering and signal enhancement will be accomplished through the combination of the well-known linear prediction filtering (LPF) and a new signal enhancement method that is based on time synchronous averaging (TSA) technique. The purpose of the LPF is the filtering of deterministic noise while the purpose of the proposed technique is to optimize TSA process by automatically aligns bearing signal time segments for best bearing slippage effect compensation.

5. **1 Envelope window selection**

Amongst several methods that are used for envelope window selection and were previously discussed in Section 2.2, the SK-based methods have shown a good capability in finding the frequency bands that contain the highest signature of faulty bearing signal. This capability returns to the sensitivity of the SK to the signal non-stationary property which is a core discriminator of transients generated by rolling element bearing faults from other deterministic transients included in the signal. The definition of the SK was earlier defined by Dwyre (Dwyer, 1983) for non-Gaussian signal detection purpose. Although several trials for defining and using the SK were addressed, such as, by Pagnan and Otonnello (Otonnello, et al., 1994) and Capedevielle et al. (Capdevielle, et al., September 10-13, 1996), a clear formal definition of the SK, especially for the non-stationary signal, delayed until the work done by Antoni (Antoni, 2006). According to the definition of the SK it’s the calculated fourth statistical moment for the complex envelope of the signal that can be produced by any of the several time-frequency domain transformation methods such as short time Fourier transform (STFT), wavelet transform (WT) or Wigner-Ville transform (WVT). The application of SK for bearing fault detection using all of the three pre-mentioned methods was studied and compared by Sawalhi et al. (Sawalhi, et al., 2004). According to the last mentioned work the STFT-based SK gives good results and less complexity than the other two methods. However, the choice of window length and overlapping ratio is a concern when using STFT-based SK for envelope window selection. The choice of STFT window length to use for SK calculation is restricted by two important facts: firstly, the STFT is an averaging process as it includes the approximation of the Fourier transformation (FT) process, as explained by Equation (5.1) below:

$$X(f,W) = \int_{-\infty}^{\infty} x_w(t)\exp(-2\pi f\,\mathrm{i})\mathrm{dt} \tag{5.1}$$

where $x_w(t)$ is the time domain vibration signal in a specified window w, $X(f, W)$ is the Fourier transform at frequency f inside the window w that has a window length W.

The result of the above definition of STFT process assists the need for longer windows to achieve better spectral resolution. In other words, increasing the number of spectral lines that cover the whole spectrum according to Shannon theorem (Proakis, et al., 1996). The application of this condition to calculate the minimum windows length requires the pre-knowledge of the highest expected frequency in addition to the sampling frequency. For example, if a sampling frequency $F_s =$ 20 KHz and a maximum expected frequency, $f_d = 500$ Hz are assumed then the number of the signal sample points that represent the distance between two pulses of that highest frequency is $W_d = \frac{F_s}{f_d} = \frac{20000}{500} = 40$ samples. Hence, the minimum number of samples in each window, according to Shannon theorem is, $W_{\min} = 2.5 \times W_d = 2.5 \times 40 = 100$ samples.

Secondly, based on SK formal definitions presented in (Antoni, 2006) and formulated by Equation (5.2) and its relation to the window length as formulated by Equation (5.3), below, the shorter the STFT window the larger the number of windows that the whole acquired signal can be divided to and hence the higher the resulted value of the SK.

$$K(f,W) = \frac{mean\left(X^4\left(f, W_{1\to N_W}\right)\right)}{\left(mean\left(X^2\left(f, W_{1\to N_W}\right)\right)\right)^2} - 2 \tag{5.2}$$

$$K_{N_w}(f) = fn\left(\frac{1}{W}\right) \tag{5.3}$$

In the above two equations, K (f, W) is the spectral kurtosis value at certain frequency f when using the window length of W samples and the resulted number of windows is N_w.

According to the previous two conditions, the length and overlapping ratio of the STFT window should be optimized such that it is short enough for high SK value and long enough for a minimum spectral resolution requirement.

In the current study for the choice of optimum window length for the STFT-based SK method, a computational algorithm that iteratively changes the window length and calculates the SK value for each window length while keeping a constant overlapping ratio is proposed. The constant overlapping ratio assumption is based on the results of the study done by Wang et al. (Wang, et al., 2011) where it was shown that, increasing the overlapping ratio till reaching some ratio achieves the least spectral leakage through the maximization of spectral side lobes damping ratio and the decrement of the main lobe spectral line width. Also, the results of studying the effect of overlapping ratio on both SK value and the spectral resolution by (Antoni, 2006; Wang, et al., 2011) showed that an overlapping ratio greater than 0.75 is recommended. Although these studies also showed that increasing of the overlapping ratio higher than the 0.75 enhances the spectral resolution, this enhancement is very small and does not worth the increase of the computational cost. Hence, a constant overlapping ratio of 0.75 will be used in the proposed window size selection computational algorithm.

Figure 5.1 shows a flowchart for the STFT-based SK estimation process including the window length selection iteration process. The input data is the discrete signal $x(n)$ that represents a faulty bearing signal. $W_{\min}$ is the minimum window length that is calculated according to the first condition explained in the previous paragraph and Δ is the window length increment by each iteration step. The signal $x(n)$ is divided into a number of windows or data segments that have a fixed length, W and have an overlapping ratio equals 0.75. Then the STFT method is used to transform the segmented signal into the time-frequency domain and the kurtosis is calculated for each spectral line over all of the windows. The maximum SK at the used window length and the adjacent frequency are stored to be compared by the results of different window lengths generated in the iterated steps. The steps that include window length changing and maximum SK calculation are repeated after increasing the window length by 60 samples, 60% of minimum window length, for each step. The iteration is repeated till the SK maximum value is almost stable or the frequency at which the max SK occurs is consistent. The size of the window that produces the largest value for the SK is chosen to be the optimum window length and the frequency at which the maximum SK achieved is chosen to be the band-pass filter center frequency while filter band width will be selected to be 20 times of the shaft frequency as recommended in many practical references of high frequency resonance, HFR, method application, e.g., (McFadden, et al., 1984; Bechhoefer, et al., September 27-October 1, 2009).

5.1.1 Application of EW selection based on STFT-SK method

The bearing fault characteristic frequencies can be calculated if the bearing dimensions and the operating speed are known. The equations that calculate bearing faults characteristicfrequencies are given in many textbooks and papers such as (Randall, 2011; Balderston, 1969; Ho, et al., 2000). These equations are represented below:

$$\begin{aligned} BPFI &= \frac{n \times f_r}{2}\left(1 + \frac{d_m}{D_b}\cos\varphi\right) \\ BPFO &= \frac{n \times f_r}{2}\left(1 - \frac{d_m}{D_b}\cos\varphi\right). \\ FC &= \frac{f_r}{2}\left(1 - \frac{d_m}{D_b}\cos\varphi\right) \end{aligned} \tag{5.4}$$

where *BPFI* is the ball pass frequency over inner race fault, *BPFO* is the ball pass frequency over the outer race fault, *FC* is the cage frequency, f_r is the shaft rotational frequency and φ is the bearing contact angle. The samples of the raw faulty bearing signals are acquired using the bearing test rig that was described previously in Section 4.3. Figure 5.2 shows the raw signal for an inner race faulty bearing when the shaft speed is 100 Hz and the sampling frequency is 20 KHz

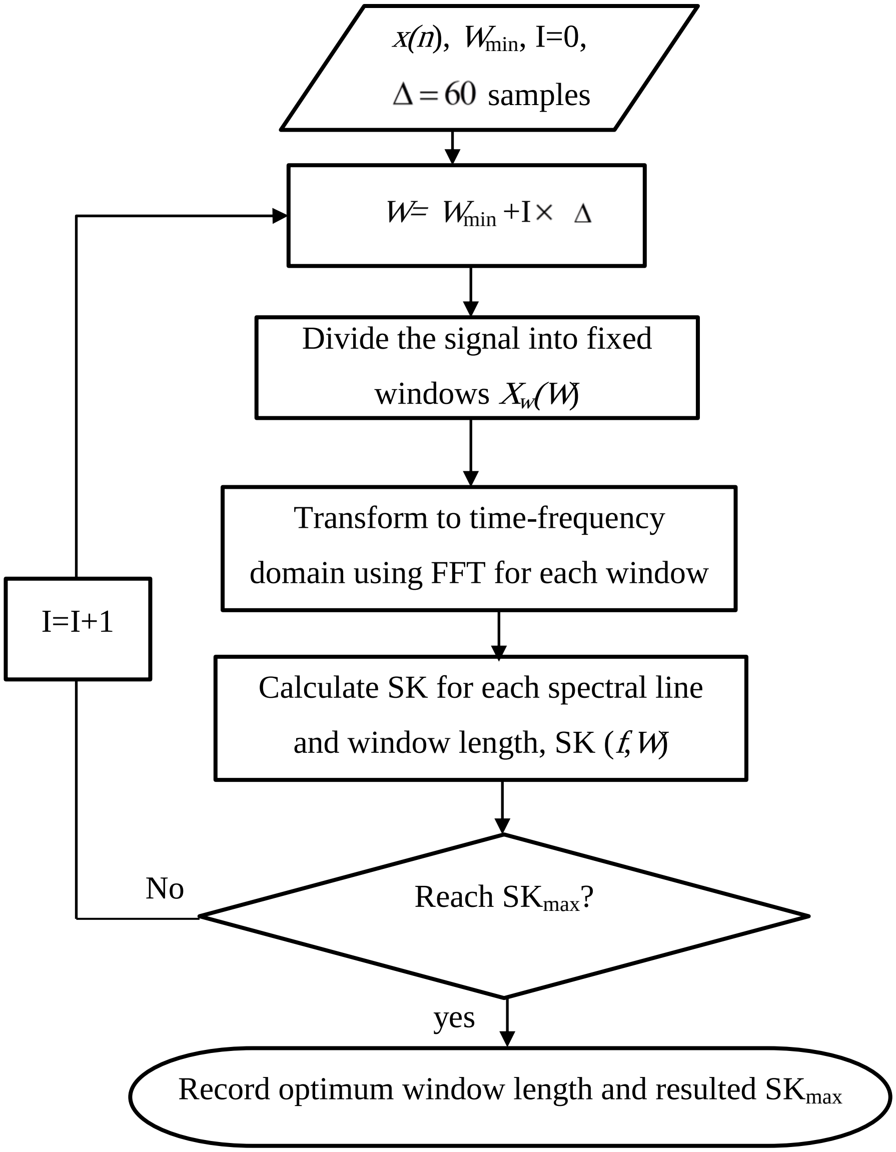


Figure 5.1 Window length selection algorithm for STFT-based SK method.

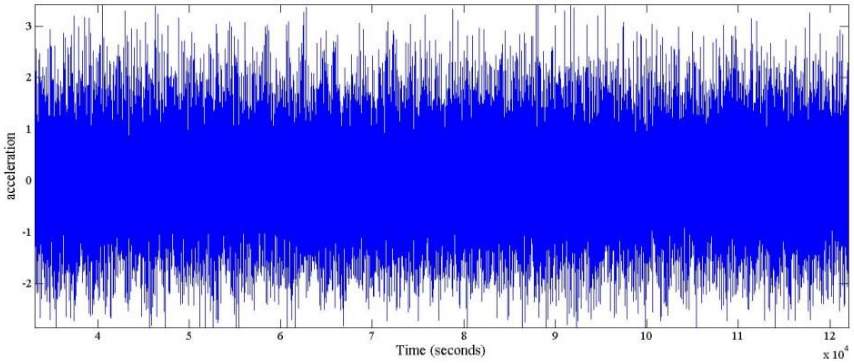


Figure 5.2 Raw signal acquired by an accelerometer for a bearing with an outer race fault running at steady speed of 6000 rpm

Applying the window selection algorithm, described in the previous section, the optimum envelope window is selected for the signal of an outer race faulty bearing that has a shaft speed of 100 Hz through the following steps: (1) calculate the fault characteristic frequency using Equation (5.4) and then the minimum window size is calculated according to the definition of minimum window length from the previous section. (2) transform the raw signal from time domain to time-frequency domain using the STFT method while keeping a constant window overlapping ratio, 0.75, and changing the window length starting from the minimum window length as calculated from the previous section (100 samples) and using an increment step of 60 samples, Figure 5.3 shows a representation of the signal as transformed to time frequency domain by the STFT method. (3) The SK value for each window length step is calculated and a two-dimensional array that stores the SK values and adjacent frequencies for each window length is established, Figure 5.4 shows a representation of SK values and the adjacent frequencies at some selected windows of different size.

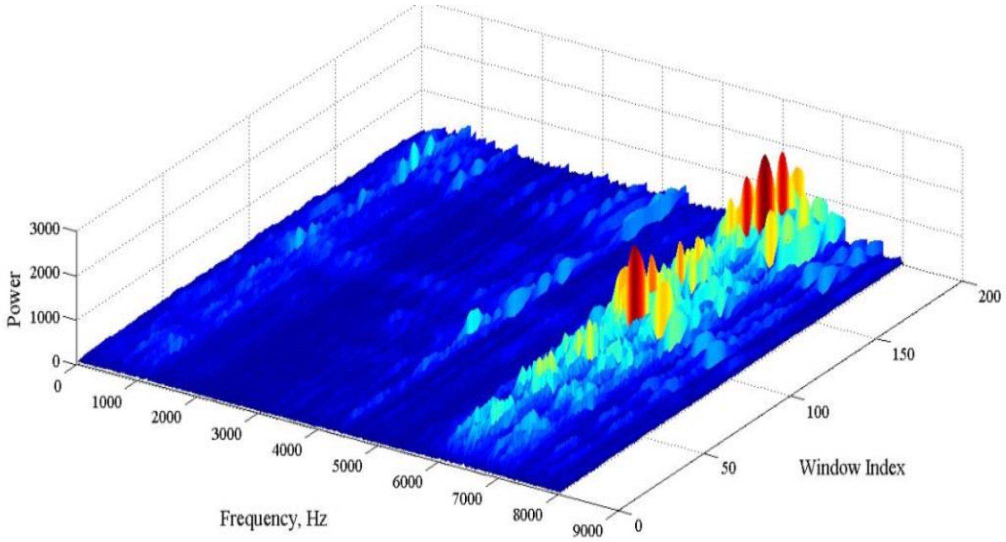


Figure 5.3 The time-frequency representation of the bearing signal obtained using the STFT method

(4) obtain the maximum value for the SK calculated for each window length and specify the frequency at which the maximum SK occurred then use this frequency as the envelope window (the bandpass filter) center frequency. (5) Set the bandpass bandwidth to equal 20 times the shaft speed frequency, i.e, for 100 Hz shaft speed the bandpass filter will have a 2 KHz band width.

Figure 5.4 shows that the SK values at different frequencies change with respect to the window length. It can be noticed that the SK values are very low at almost all frequencies when using a window length smaller than the calculated minimum window length. This is the case when the window length is only 60 samples while the previously calculated minimum window length is 100 samples. The figure also shows that the SK values increase by increasing thewindow length till reaching a 280 samples window length. Increasing the window length more than the 280 sample window does not increase the SK values. Thence, the optimum window length is selected to be 280 samples and the envelope window central frequency is chosen to be the frequency at which the maximum SK was found. For the case of the inner race faulty bearing the central frequency was found to be 7763 Hz as shown in Figure 5.4

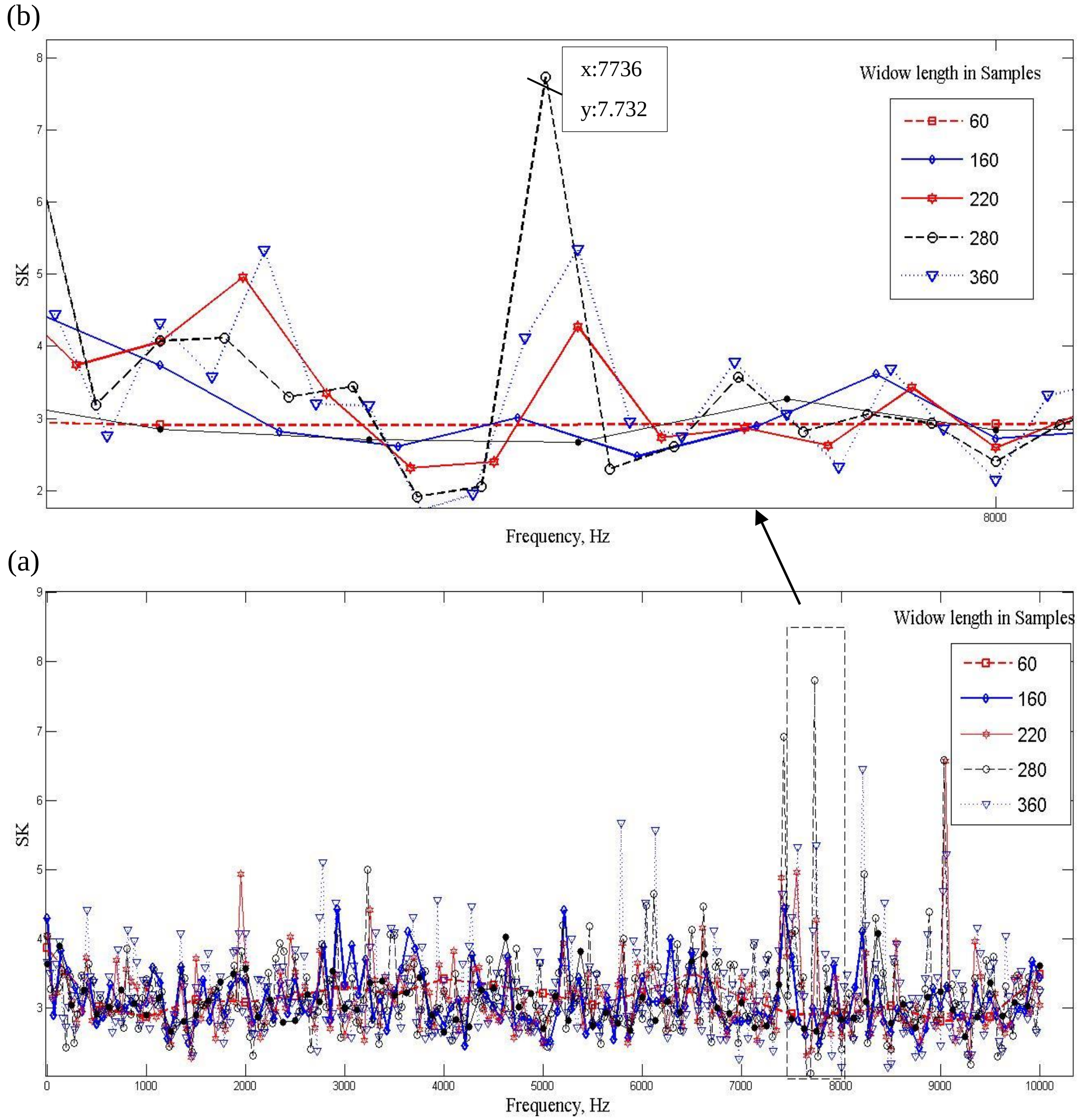


Figure 5.4 (a) The SK values as calculated at any frequency and using different window lengths, and (b) a zoomed section at maximum SK

The values for central frequency and optimum window length are calculated for each expected fault in separate proceedings as the fault characteristic frequency is involved in calculating the window length and the SK. Using the above procedures for different bearing health conditions and for the same shaft speed, 100 Hz, the optimum central frequencies for each case and the bandpass filter lower cutoff, f_{cl} and upper cutoff, f_{cu}, frequencies are listed below in

Table 5.1:

Bearing condition EW band pass filter parameter	Healthy bearing	Inner race fault	Outer race fault
f_c (Hz)	3246	7736	6493
f_{cl} (Hz)	2246	6736	5493
f_{cu} (Hz)	4246	8736	7493

Table 5.1 Selected EW bandpass filter parameters: center frequencies and upper and lower cut off frequencies for different bearing conditions based on the SK method (shaft frequency = 100 Hz)

5.2 In-band signal enhancement

The previous step was to select an envelope window that best represents the impulsive nature of the bearing vibration signal. However, the impulsiveness resulted from an incipient fault usually has a weak amplitude compared to other system components' vibration even if an optimum frequency band window for selected (Randall, et al., 2011). Therefore, an in-band bearing signal enhancement is needed reveal any impulsive nature due to incipient faults.

This will be accomplished through two steps: removal of deterministic part of the signal using a linear prediction filter (LPF), and fault signal enhancement using autonomous time synchronous average (ATSA). Despite the time synchronous averaging method has been largely used and developed for decades as a powerful method to enhance the periodic component from mixed components signal (Braun, 2011), it was mostly successful for gear fault detection (McFadden, et al., 2000). However, for bearing signal enhancement this method faces the problem of rolling element slippage. The slippage effect is caused by the randomness of the actual time of ball passing frequency and the deviation of its value from that values calculated by bearing kinematic equations. Although the percentage of slippage in rolling element bearing is usually small, if compared to the shaft cycle time, it converts the bearing signal from stationary to cyclostationary (Antoni, et al., 2004). In addition to the distortion of the envelope spectrum caused by the slippage effect (Sawalhi, et al., 2007), the cyclostationary nature of the bearing signal can affect the performance of the signal enhancement methods that rely on the averaging of the periodic part of the signal. This problem can be partially overcome for very low slippage percentage if the start and the end of the cycle events are well defined using an angular position measurement device. Although, the angular measurement device can correct the effect of the slippage between the bearing races and the driving element, either the shaft or the housing, it cannot compensate for the slippage between the bearing internal parts. Also, as a result of the inaccuracies of synchronizing device and the unavoidable shaft speed variations affect the efficiency of the synchronous averaging method. The degradation of synchronous averaging method results can be more obvious in the case when the operating conditions featuring lightly loaded bearing with high shaft speed (Harris, et al., 2007), as these are the most likely conditions for high slippage ratio.

The purpose of the proposed ATSA method is to exploit the cyclostationary feature of bearing fault signal to re-align the signal such that the bearing fault signal is converted from random to deterministic signal, hence improving the efficiency of the averaging process.

So the current section is working on the in-band enhancement of faulty bearing signal through two steps: the first is the removing of the deterministic noise such as shaft and gear meshing harmonics and the second is the strengthening of the bearing weak signal by a new modified synchronous time averaging method, the ATSA.

5.2.1 Deterministic signal removal using a linear prediction filter

The first step in the scheme is to remove the deterministic part from the bearing signal. This is done using an adaptive length linear prediction filter. The filter estimates the current values of the signal based on the immediate history of the signal. This is formulated as follows:

$$\hat{x}(n)=\sum_{k=1}^{p}a(k)x(n-k) \tag{5.5}$$

where $x(n)$ is the time signal, $\hat{x}(n)$ is the estimated signal and $a(k)$ is the filter parameters that can be calculated using Yule-Walker equation that is solved by the Levinson–Durban recursion (LDR) algorithm (Randall, et al., 2011; Proakis, et al., 1996). The order of the filter can be chosen based on minimum least square errors calculated by subtracting the estimated signal from the real signal. The residual should contain the entire non-periodic signal and should be rich in the bearing fault signal.

5.2.2 Fault signal enhanced by autonomous time synchronous averaging

A solution for extracting the bearing fault signals from a partially corrupted signal due to the slippage of the rolling element will be tested in the current section. The method concerns not only the signal de-noising, but also compensating for the possible effect of rolling element slippage to enhance the fault characteristic power with respect to the powers of other sources in the signal. This is achieved by revealing the localized impulsive nature of the bearing signal using the Teager operator (TO).

The Teager operator was firstly proposed by Teager (Teager, et al., 1990) and used for speech modeling and production applications. Since then it showed a good efficiency in separating the amplitude modulated and frequency modulated speech signal and in extracting the strongly amplitude modulated signal such as gear meshing signal (Bozchalooi, et al., 2010). The continuous and discrete versions of the Teager operator are given by the following two equations respectively:

$$TO\left[x(t)\right] = \left[\dot{x}(t)\right]^2 - x(t)\,\ddot{x}(t) \tag{5.6}$$

$$TO\left[x(n)\right] = x^2(n) - x(n+1)\,x(n-1) \tag{5.7}$$

The proposed algorithm uses the TO as an impact, bearing fault impulses, locator for the signal in the time domain. Firstly the acquired signal is divided into equal length segments. The length of each segment should represent one shaft cycle. The choice of a one shaft cycle as a length of the data segment is important to ensure that for a steady shaft operation the only non-stationarity property inside the signal should be referred to the bearing slippage. Then the TO for each sample point in the segments is calculated and only the probable fault impacts are considered while the peaks that are very close to each other are neglected. The peak point selection is constrained by a minimum spacing condition (MSC) that is defined as the minimum acceptable spacing between any two peaks and between any peak and the boundary points of the segment, i.e. first and last points. The value of MSC is in samples and can be calculated for each expected fault frequency f_f and sampling frequency F_s as follows:

$$MSC = \frac{F_s}{f_f} \pm 10\% \times \frac{F_s}{f_r} \tag{5.8}$$

In the above equation f_r is the shaft rotational frequency while the minimum spacing between two fault impacts should not be less than the spacing calculated by the ideal f_f, calculated by Equation (5.4) and with an acceptable error that should not exceed 10% of the spacing representing one shaft cycle.

According to the pattern of the TO peaks that are selected for each data segment and the comparison to the possible fault frequencies, the bearing fault is identified and the segment with most ideal spacing between the TO peaks, regarding the time between fault impacts, is selected as the reference segment. For each of the other segments the location of the TO peaks is compared to the reference segment and the points that have TO peaks are shifted locally to be aligned with the adjacent peak points in the reference segment. Assuming that the maximum rolling element slippage is in the range of ± 2% of the shaft speed then the maximum allowable shifting (MAS), in samples, can be calculated as follows;

$$\mathrm{MAS} = \frac{F_s}{f_r} \times \left(\pm \frac{2}{100} \right) \tag{5.9}$$

where F_s is the sampling frequency and f_r is the shaft rotational frequency

After accomplishing the alignment process for the entire peak points a new set of data segments results. The time averaging is accomplished for both original data and aligned data the last step is to predict the envelope spectrum for both the original averaged and the aligned averaged signals and compare their values at the fault characteristic frequencies. The steps for the proposed algorithm are listed below:

1. Divide the sample of the raw signal into N equal length windows, the length of the window should represent one shaft revolution or an integer number of shaft revolutions, depending on the sample length and the sampling frequency. Hence, each window length as a number of sample points, W, can be calculated depending on the sampling frequency, F_s, and the shaft speed, f_r. For example for a shaft speed, f_r =100 Hz, and a sampling frequency, F_s=60 KHz, the window length is calculated as follows.

$$W = \frac{1}{f_r} \times F_s = \frac{1}{100} \times 60000 = 600 \ \text{samples} \tag{5.10}$$

2. For each window, calculate the Teager operator value TO[$x(n)$] using Equation (5.7) for each data point, and select the highest values in each window that validate the MSC condition as in Equation (5.8). For example, the number of TO peaks can be from 3 or 5 peaks in the shaft revolution segment depending on the fault frequency. The fault type is detected by comparing the number of peaks in each revolution that satisfy the MSC condition.
3. Select the reference segment that has the best impulsiveness feature by checking that the spacing between any two fault peaks, PS, is the closest to the ideal spacing between two consecutive fault impacts. This condition for the reference segment can be formulated as follows:

$$N_{ref} = \arg_{PS}\left[\min\left(Ps - \frac{F_s}{f_f}\right)\right] \tag{5.11}$$

4. For each of the other data segments compare the local position of each segment peak with the adjacent peak in the reference segment. Then relocate the peaks in each segment by doing step by step shifting such that the maximum number of samples of the shift is less than MAS calculated by Equation (5.9). For example, for a 100 Hz shaft speed and sampling frequency of 60 KHz, the MAS should not exceed 12 samples which means that the window can be shifted by ± 6 samples.
5. Perform time averaging for a synthesized data that is built from the aligned segments and calculate the envelope spectrum for both original and synthesized data.

5.3 Validation of the proposed algorithm for bearing fault detection

The discussed methods in the previous two sections are illustrated by the flow chart in Figure 5.5 and then are validated using both simulated and experimental data in the next two sub-sections:

5.3.1 Validation using simulated signal

Faulty bearing vibration signal has been defined and simulated in many research works such as (McFadden, et al., 1984; Ho, et al., 2000; Bozchalooi, et al., 2007). The simplest model for bearing fault will consider the fault signal as an impulse signal occurring at nearly periodic intervals and exciting the structure resonance frequency that decays due to the structure damping. In other words it can be assumed that a fault signal which

has impulsive nature modulates the structure natural frequency, considered here as the carrier signal, and the modulated signal that decays according to the damping character frequency. The fault signal model incorporating the above is represented by the following equation:

$$x(n)=\sum_{m=1}^{M}\Big(\exp\big(-\beta\left(n\text{-}m.T_i\right)\big)\cdot\cos\ \big(2.\pi.w_r.\left(n\text{-}m.T_i\right)\big)\cdot \mathrm{Dirac}\big(n\text{-}m.T_i\big)\Big) \tag{5.12}$$

where w_r is the natural frequency of the structure, β is the structural damping characteristic frequency, m is an index of the fault impacts, T_i is the time period between two consecutive impacts, and M is the total number of impacts expected of the whole sampling interval. The simulated fault signal using the above equation for slow and fast shaft speeds are shown in Figure 5.6. The main difference in the two simulated faulty signals in Figure 5.6 is the effect of the shaft speed that will cause the fault impulses to overlap for fast shaft while for slow shaft the fault impulses will completely decay and vanish before the occurrence of a subsequent impact.

5.3.1.1 Slippage effect removal.

The bearing signal randomness due to rolling element slippage can be simulated by changing the impact period T_i in Equation 5.6 randomly such that the impact time is no more deterministic. This can be accomplished using a random number generator and set the randomness limits of the rolling element impact time between either $\pm 2\%$ for low to moderate slippage conditions and $\pm 5\%$ for high slippage condition. By adding the slippage effect to the faulty bearing simulated signal a more realistic simulation that has a non-uniform distance between any two consecutive impacts is approached. The simulated signal while considering possible slippage is depicted in

Figure 5.7.

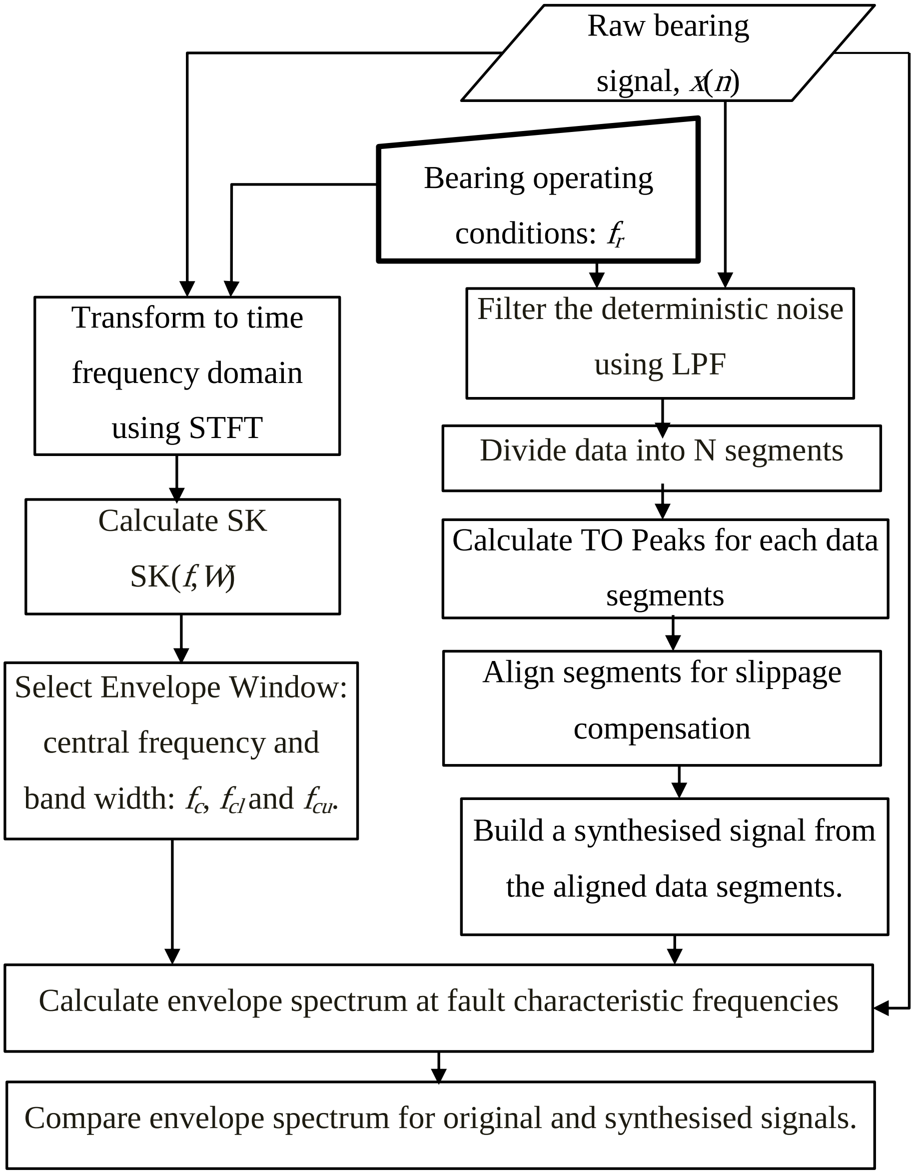


Figure 5.5 Vibration-based bearing fault detection scheme.

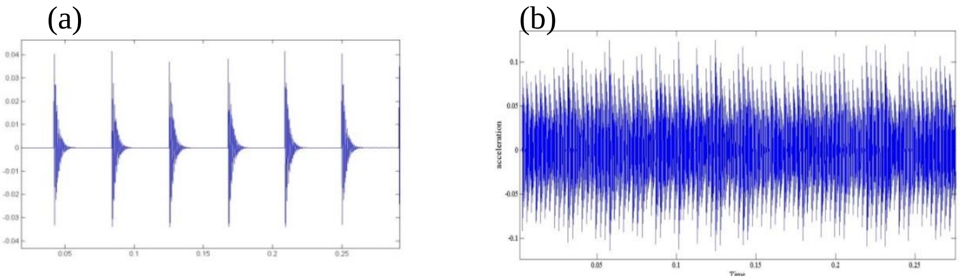


Figure 5.6 Simulated bearing fault at different shaft speeds: a) at 8 Hz, and b) at 100 Hz.

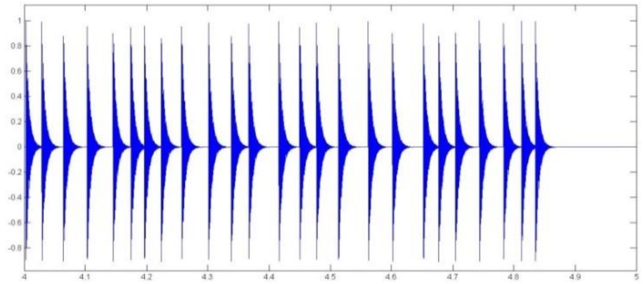


Figure 5.7 Simulated bearing fault with the slippage effect.

It can be shown that the bearing slippage has an effect on the envelope spectrum even when the signal is monotonic and zero noise is considered. This result is depicted in Figure 5.6 a-c in which a clean monotonic bearing fault signal is simulated with different slippage ratio. For a zero slippage ratio as in Figure 5.6(a) the envelope spectrum depicts clear peaks at the fault characteristic frequency, 346 Hz, and its higher harmonics. The slippage effect changes the fault clear peaks at Figure 5.8 (b) and 5.8(c) and flattens the peaks around the fault characteristic frequency and its harmonics. The great effect of slippage on envelope analysis technique is revealed when a noise is added to the simulated signal. Even the simplest type of noise, the additive white noise, can cause a great degradation of the envelope spectrum results for bearing signal. The results of adding a 1 dBw white Gaussian noise to the simulated signal Figure 5.8(a-c) are depicted in the same figure d-f. these three sub-figures show that the increasing the slippage ratio lead to the complete disappearing of the fault signal peaks especially when the slippage ratio is above 2% as in Figure 5.8(f). Both the TSA and the proposed ATSA methods are applied to enhance the degraded envelope analysis due to bearing signal slippage effect. The envelope spectrum results obtained after enhancing using TSA are shown in Figure 5.8 (g-i) while the envelope spectrum results obtained after enhancing using the ATSA method are depicted in Figure 5.8 (j-l). It can be seen that the ATSA was successful to reveal the fault signal peaks at the two cases with different slippage ratios while using the TSA could not enhance the envelope analysis efficiently enough to reveal the fault signal peaks. The comparison between using the two methods, TSA and ATSA, for the enhancement of envelope analysis method can be clarified by comparison of the signal to noise ratio (SNR) resulted by each method and for each assumed slippage ratio. The lists of resulted SNR for each case that are depicted in Figure 5.8 are tabulated in the following table:

Slippage Ratio Analysis method	0	± 2%	± 5%
Envelope analysis (EA) alone.	15 dB Figure 5.8 (d)	1.7 dB Figure 5.8 (e)	-20 dB Figure 5.8 (f)
TSA+ EA	23 dB Figure 5.8 (g)	6.3 dB Figure 5.8 (h)	-7.2 dB Figure 5.8 (i)
ATSA+EA	21 dB Figure 5.8 (j)	14 dB Figure 5.8 (k)	7.8 dB Figure 5.8 (l)

Table 5.2 Envelope spectrum enhancement, increase of SNR, for different bearing slippage ratios as depicted in Figure 5.8 (d-l).

5.3.1.2 Discrete noise removal.

Usually the bearing signal is very vulnerable to be masked with other system components vibration signals that are often called interferences. One of the most probable sources of this type of interference is the signal generated by gear meshing. This type of faults has higher power than the bearing signal and usually has a higher frequency, several order of shaft frequency. In current sub-section this type of interference is simulated and added to the bearing simulated signal then a

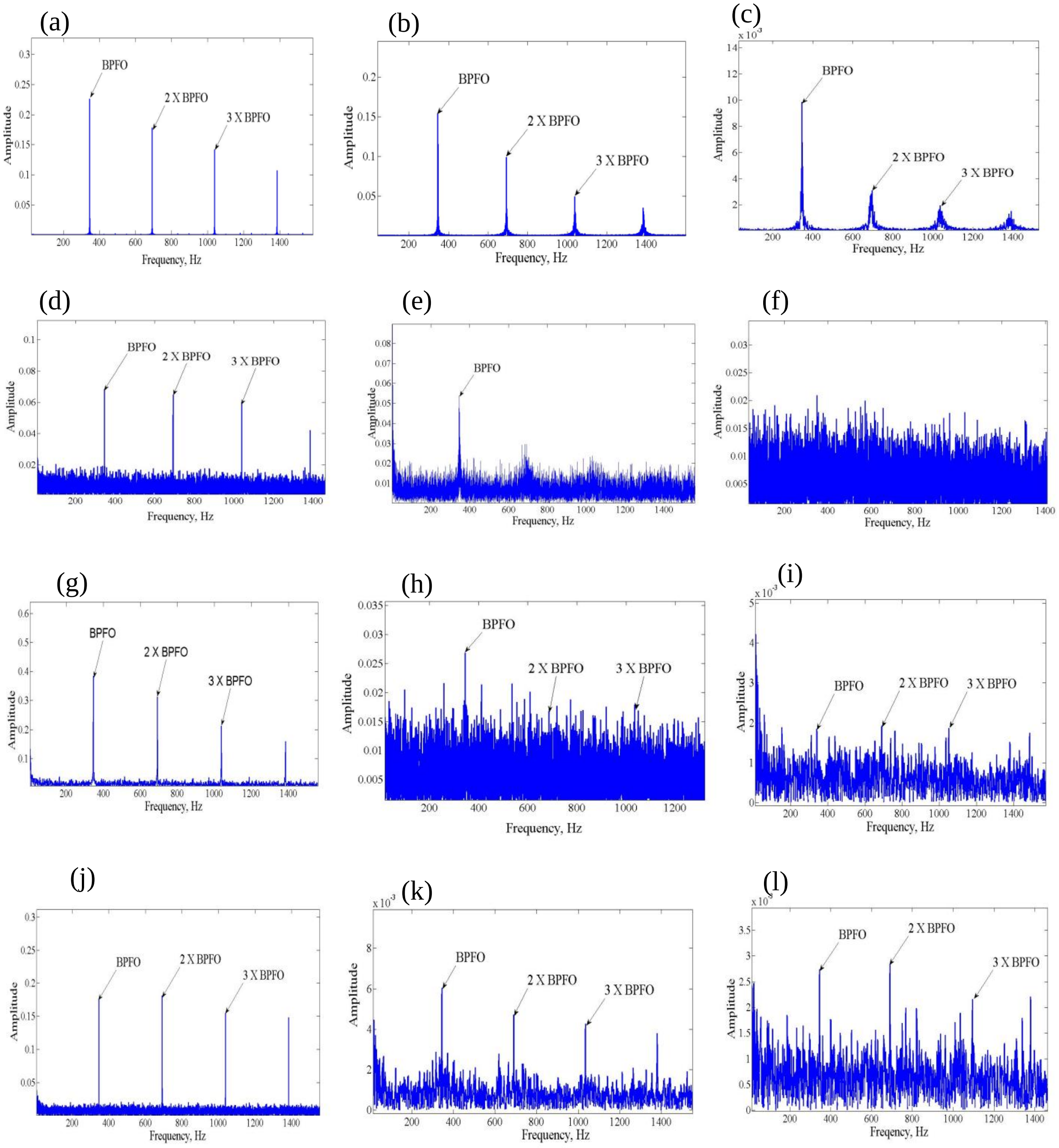


Figure 5.8 Slippage effect on bearing envelope spectrum appeared in the envelope analysis results (a-f), enhanced envelope analysis results by the TSA method (g-I) and enhanced envelope analysis by the ATSA method (j-l). Knowing that slippage ratio is zero for a, d, g and j, 2% for b, e, h and k and 5% for c, f, i and l, the first raw represents the effect of the slippage on a clean signal, i.e. zero noise while for all other simulated signals a white Gaussian noise of 1 dB is added.

combined LPF and either TSA or ATSA are used and compared to the envelope spectrum that is calculated for the simulated signal without these methods. The model for the signal that contain both bearing and gear signals can be expressed as follows:

$$z(n) = x(n) + y(n) \tag{5.13}$$

where $z(n)$ is the modeled signal that contains both bearing and gear vibration signals, $x(n)$ and $y(n)$, respectively. The model of the bearing signal was presented by Equation (5.12) while the model for the gear meshing signal is represented by the following equation:

$$y\left(n\right) = \sum_{k=1}^{K} \left(\exp\left(-\beta\left(n\text{-}k.T_g\right)\right) \cdot \cos\left(2.\pi.w_r.\left(n\text{-}k.T_g\right)\right) \cdot \mathrm{Dirac}\left(n\text{-}k.T_g\right)\right) \tag{5.14}$$

where w_r is the natural frequency of the structure, β is the structural damping characteristic frequency, k is an index of the gear meshing impacts, T_g is the time period between two consecutive gear meshing, and K is the total number of impacts expected of the whole sampling interval. The simulated signals that represent the interference of bearing and gear signals are shown in Figure 5.9(a and b) when the slippage ratio is ± 2% and ±5%, respectively. Assuming that the bearing signal is the targeted signal and the gear meshing signal is an interference that is required to be removed, the signal to interference ratio (SIR) will be considered as a good measure of the efficiency of deterministic noise removal methods. The signal to interference ratio is defined as the ratio of the signal power to the interfering signal power. In case of using the SIR to compare the enhancement of the envelope spectrum using different methods, the power of the envelope at the signal characteristic frequency, bearing fault characteristic frequency, and its harmonics is compared to the power of the interfering signal, gear meshing signal, and its harmonic in certain range of frequencies. According to the previous definition the SIR calculated for certain frequency band can be calculated using the following equation:

$$SIR_{f_1 \to f_2} = 20 \times \log_{10} \left(\frac{\sum_{i=1}^{n_b} S\left(f_i\right)}{\sum_{j=1}^{n_g} I\left(f_j\right)} \right) \qquad (5.15)$$

In the above equation the SIR is calculated for the frequency band starting at frequency f_1 and ends at frequency f_2, $S(f)$ is the power of the bearing signal at certain frequency f while i is the index for the bearing signal harmonics and n_b is the number of bearing signal harmonics in the SIR calculation frequency band. Similarly, $I(f)$ is the power of the interfering signal while j is the index for the gear meshing signal harmonics and n_g is the number of gear meshing signal harmonics in the SIR calculation frequency band.

Using the above definition, the SIR for the envelope spectrums calculated by the basic envelope analysis alone and shown in Figure 5.9 (c) for the bearing signal with ± 2% slippage and in Figure 5.9(d) for the bearing signal with ± 5% slippage are -21 dB and -53 dB, respectively. It can be said that the gear meshing vibration signal dominates the envelope spectrum for both levels of bearing slippage and for the higher slippage ratio, ± 5%, the bearing signal is completely covered by noise and only interfering signal is revealed. Then both LPF and TSA methods are applied on the signal before envelope analysis. The reconstructed signals after TSA for bearing signal with the two slippage ratios are depicted in Figure 5.9 (e and f), respectively. The envelope spectrums resulted from the combined LPF and TSA are shown in Figures 5.9 (g and h) where it can be seen that the bearing signal is revealed when the slippage ratio is still low, Figure 5.9 (g), with SIR equals 11 dB while it is only 1.45 dB for the higher bearing slippage ratio, Figure 5.9 (h). Figures 5.9 (i and j) shows the two slippage ratios signal after filtered by LPF method and then reconstructed by the ATSA method. The effect of pre-alignment of the signal before executing the time averaging, the ATSA method, produces a higher power of the reconstructed signal at the bearing fault frequency which leads to a higher SIR for the envelope spectrums shown in Figures 5.9 (k and l). The calculated SIR for the envelope spectrum in Figure 5.9 (k) is 32 dB while it is only 20 dB for the signal with the higher slippage ratio, Figure 5.9 (l). The results of envelope analysis enhancement through interference removal that are depicted in Figure 5.9 are listed in Table 5.3 below:

Slippage Ratio / Analysis method	± 2%	± 5%
EA alone.	-21 dB SIR, Figure 5.9 (c)	-53 dB SIR, Figure 5.9 (d)
LPF+TSA+ EA	11 dB SIR, Figure 5.9 (g)	1.45 dB SIR, Figure 5.9 (h)
LPF+ATSA+EA	32 dB SIR, Figure 5.9 (k)	20 dB SIR, Figure 5.9 (l)

Table 5.3 Enhancement of SIR for the envelope spectrums expected using three different methods and at two different ratios of bearing slippage

It can be concluded from the results of the two subsections that bearing slippage affects the detection of bearing fault vibration signal either in the presence or the absence of interfering deterministic noise. It is obvious also that using the envelope analysis alone to detect bearing faults when the bearing slippage is expected result in very poor detection capability. On the other hand preceding the envelope analysis with either LPF and TSA or LPF and ATSA was successful to enhance envelope analysis as a bearing fault detector. The superiority of the ATSA method over the TSA method can be concluded from the higher SNR and SIR enhancement when using ATSA as shown in Tables 5.2 and 5.3.

5.3.2 Validation using experimental data

The acquired signal from the test rig described in Chapter 4 and the different bearing faults that are depicted in the next chapter will be used to test the vibration analysis methods proposed in this chapter. The test is accomplished using a bearing with an inner race fault that has a fault characteristic frequency of 486.5 Hz calculated by Equation (5.1) when the shaft speed is 100

Hz. The enhanced envelope analysis by the technique described in section 5.2.1, envelope window selection, and section 5.2.2, in-band signal enhancement, will be carried out.

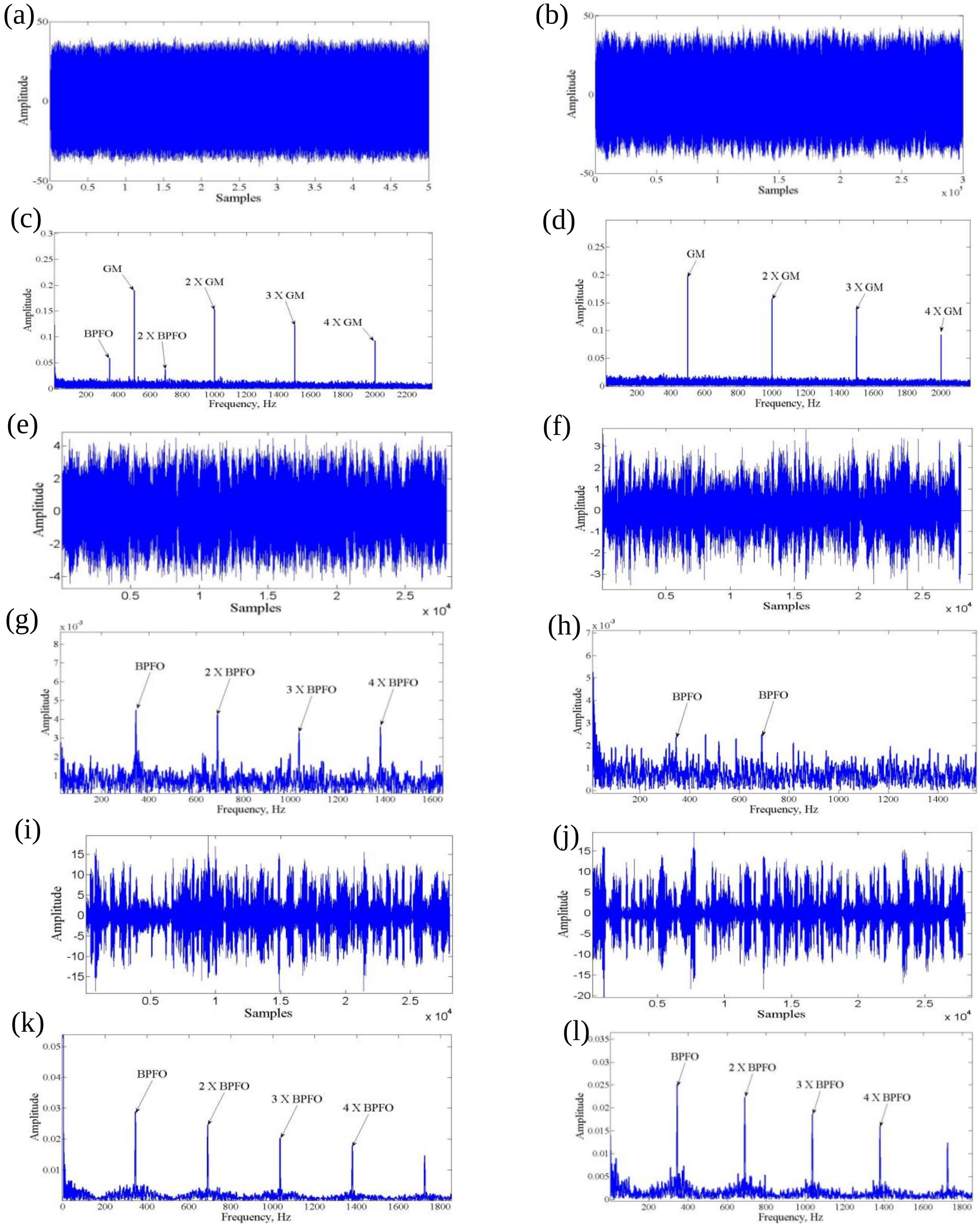


Figure 5.9 Envelope spectrums for the simulated interfered bearing signal with a simulated gear meshing signal when considering a ± 2% bearing slippage (a) and ± 5% bearing slippage (b). (c) and (d) represent the envelope spectrums by the basic envelope analysis method, (e) and (f) represent the reconstructed signals using the TSA method while (g) and (h) are the envelope spectrums for the reconstructed signals (e) and (f), respectively. Plots (i) and (j) represent the reconstructed signal using ATSA method while (k) and (l) are the envelope spectrums for the reconstructed signals (i) and (j), respectively

Firstly, the basic envelope spectrum at the fault frequency when using the envelope window selected according to the method presented in Section 5.2.1 is accomplished. The SNR of the resulted envelope spectrum will be used as a reference to be compared to the SNR of the spectrum obtained using the envelope analysis enhanced by the ATSA method. The raw signal of the inner race faulty bearing is shown in Figure 5.10. The signal is divided into a number of windows such that the windows length depends on the sampling frequency and the fault characteristic frequency as described in Equation (5.10). Then the Teager operator at each point inside each window is calculated using Equation (5.7). The two conditions of MAS and MSC, Equations (5.8) and (5.9) are applied to select most probable peaks to represent the fault impact then condition of the reference window is applied, Equation (5.11). All the windows are compared to the reference window from the peaks localization point of view. For each window single peak, a shifting process is applied till the beat alignment of each single peak with its adjacent peak in the reference window. The shifting process for each single peak is limited by a maximum number of shifting in sample points that is calculated by Equation (5.9). The result of the alignment process that precedes averaging is depicted in Figure 5.11 which depicts the windows that build the raw signal before and after the alignment process that is based on the TO peaks local position.

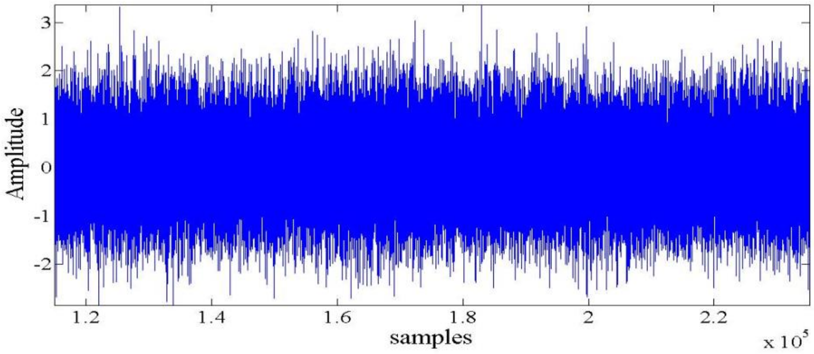


Figure 5.10 Inner race faulty bearing raw vibration signal while shaft speed is 100 Hz.

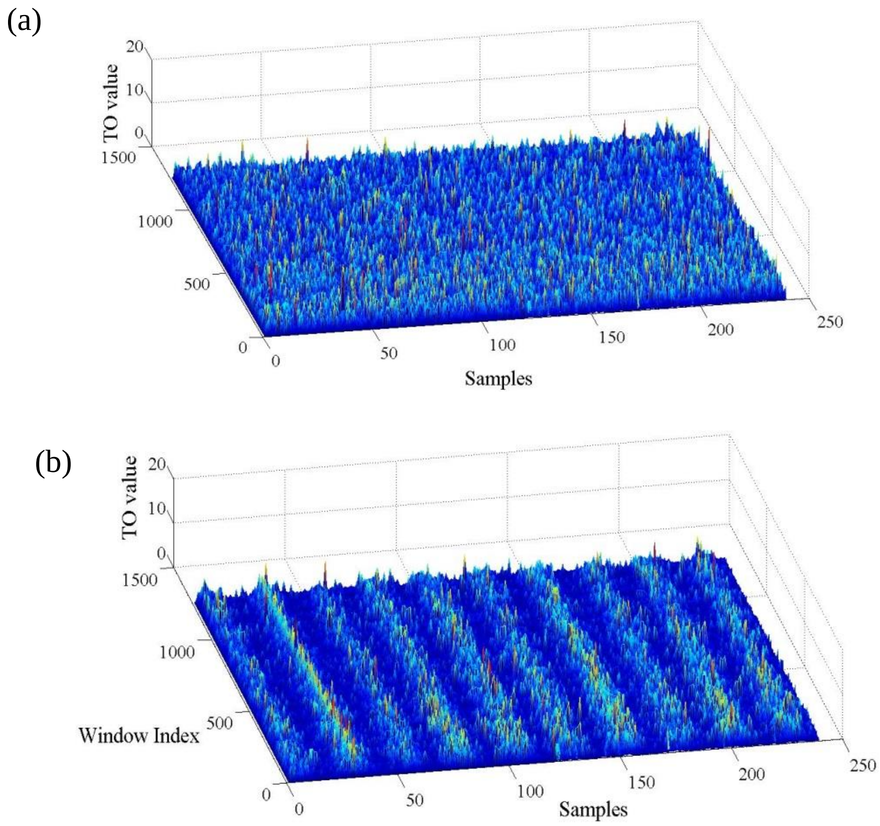


Figure 5.11 Bearing signal alignments using the TO peak localization: (a) the windows of the raw signal before alignment, and (b) the aligned windows.

Figure 5.11 shows the alignment process that divides the raw signal into 1250 data segment each segment contains 240 sample points. After the alignment step is accomplished every ten segments of the aligned signal, Figure 5.11 (b) are averaged and the resulted averages segments, 125 averaged data segments are used to build a new data array which of length equals 240×125 sample points. The envelope spectrums for the new reconstructed signal and the original signal are calculated based on basic envelope analysis and the envelope window bandpass filter that was previously selected in Section 5.1.1. The resulted envelope spectrums for both signals are shown in Figure 5.12. The figure shows that enhancing the faulty bearing signal using the ATSA method before the envelope analysis step results in enhancement of the SNR which is calculated and found to be 35 dB for the enhanced EA (Figure 5.12 (b)), while it is only 9 dB for the envelope spectrum of the raw signal without the enhancement step (Figure 5.12 (a)).

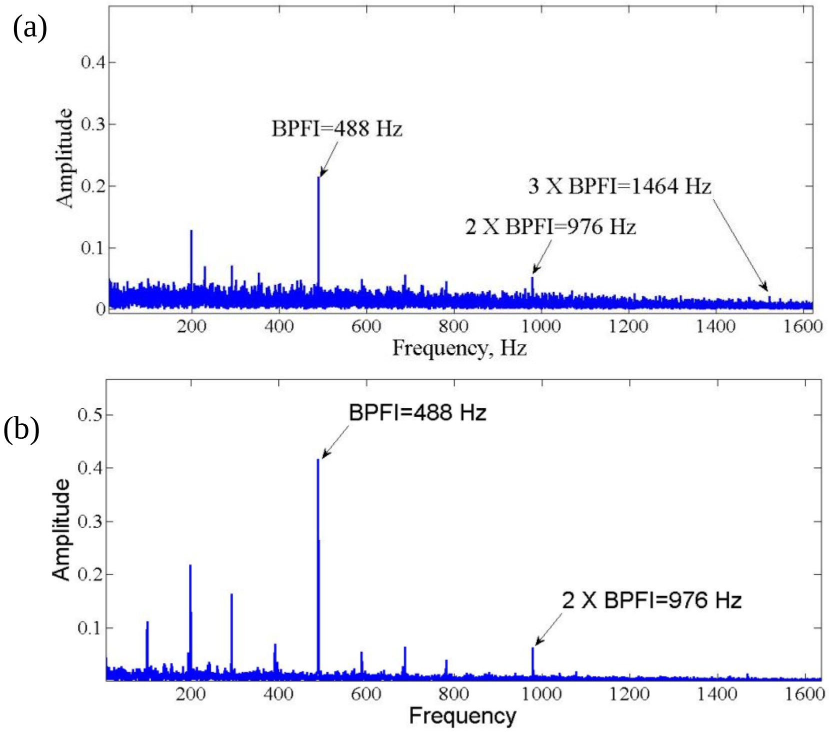


Figure 5.12 The resulting envelope spectrum for an inner race faulty bearing when using (a) the original raw signal, and (b) the signal enhanced by the ATSA method.

5.4 Summary of the bearing vibration analysis method

In the current chapter, the widely used envelope analysis is re-visited where it is shown that, for an effective EA, both proper envelope window selection and in-band signal enhancement are required. In this study, the STFT-based Spectral Kurtosis method was used to select the optimum envelope windows for different bearing faults at a certain shaft speed.

For in-band signal enhancement, two pairs of combined methods, the LPF and either TSA or the new proposed ATSA, are studied. The target of the LPF is the removal of deterministic noise while both traditional TSA and the new proposed ATSA methods are for the enhancing of the bearing signal. The new proposed ATSA is based on the time averaging technique in addition to a pre-alignment process that re-arranges a segmented copy of the raw data for best positioning of points of TO peaks.

Two obstacles that affect the EA method, the presence of deterministic noise and the rolling element slippage were simulated and tested using firstly the LPF with the TSA methods and secondly using the LPF with the ATSA methods then the before the envelope spectrums are obtained. The results showed the superiority of the ATSA method over the TSA method and in particular when there is a high rolling element slippage ratio. The application of this method on a real faulty bearing signal showed that it effectively enhances the envelope spectrum and increase the resulted SNR more than twice of the resulted envelope spectrum when the basic EA method is applied directly on the bearing raw signal while using the same bandpass filter.

Chapter 6

Bearing Fault Detection

In current chapter, both vibration and temperature methods will be used for bearing fault detection purpose. The bearing temperature calculation methods were discussed in detail in Chapter 3 while a new analytical model for bearing transient temperature calculation was presented in Chapter 4. The model that consists of three modules, dry friction calculation, hydrodynamic friction calculation and heat transfer module was tested and its results were compared with the results of two empirical methods, Palmgren and SKF models. The bearing heat generation rate calculated by the first and the second modules of the proposed model showed a good agreement with the heat generation rate calculated by both of the two empirical methods, Palmgren and SKF. The experimental test that included bearing transient temperture measurement showed that the temperature rise curves calculated by the proposed model were more accurate than the two empirical models when used with the third module for bearing temperature calculation.

An enhanced envelope analysis method that was proposed and discussed in Chapter 5 will be used in the current chapter as an example of vibration based bearing fault detection methods. The purpose of studying both vibration and temperature methods in the same research work will be revealed in current chapter as it will be explained that the pitfalls and disadvantages of each group of methods are diffrerent. In other words, the vibration based method can be good in discovering certain bearing faults when the temperature method cannot and vice versa.

The first section of this chapter will propose a new definition of the bearing faults depending on the vibration signal impulsive nature. The reason of this new categorization is to group the faults that are easier for vibration method and that are easier for temperature methods to be detected with.

The second section will present the result of the experimental work done to generate two types of faults as been defined in Section 6.1. In section 6.3 the bearing fault effect on bearing temperature is studied and proposed as a module that can be used to modify the bearing transient temperature model that is discussed in Chapter 4. Finally, several experimental tests for bearings with different faults are presented and the results are compared to the bearing temperature rise curves calculated by the analytical model and the experimental measurements for healthy bearings that are running in the same operating conditions will be discussed in Section 6.4.

6.1 Bearing faults classification

There are several different bearing faults each having a different effect on the bearing temperatures and vibration signature. To simplify the discussion about fault effects on bearing temperatures and vibration, the bearing faults are classified into two groups. The first group has an impulsive vibration signature that may be detectable over the whole bearing operation life or just detectable during a certain stage of fault maturity. The second group does not have an impulsive vibration signature and can only give a vibration indication when the fault becomes severe. The faults with impulsive vibration signatures can be further divided into two subgroups. The first subgroup features faults with a consistent vibration signature such as misalignments and imbalance faults. The second subgroup faults are those without a consistent signature such as cracks and spall faults whose vibration signature depends on the maturity degree of the fault. For example, the spall faults that usually have recognizable faulty vibration signatures (impulses) only during a certain stage of severity. At the very early stage and after a certain severity level (or size) of such faults, the faulty vibration signature cannot be easily observed (Randall, 2011). The faults of the second group that do not have impulsive vibration signature at all include the lubrication problems and incipient wear. Hence, the fault classification can be summarized as follows:

1. Impulsive faults (IF).
    a. IFC: Faults with consistently recognizable vibration signature (e.g., shaft misalignment, load imbalance, etc), and
    b. IFT: Faults with temporarily recognizable vibration signature (e.g., spall or dents in inner or outer raceways).
2. Non-impulsive faults (NIF), e.g., lubrication and incipient or gradual wear problems

The bearing faults generally fall into IFT and NIF types. Both of the two types of faults are difficult to detect over the complete range of the fault when only a vibration-based condition monitoring system is used. This means that the vibration methods can detect the bearing faults only in part of the fault range, e.g., before the fault develops to larger size or becomes blunter (Randall, 2011). When the localized fault gets blunter, the impulses of its vibration gets weak and not spiky enough to be detected as it will be buried in the system noise (Kurfess, et al., 2006). The other example of faults that the vibration methods can only detect in a limited fault range is the lubricant leak or slight deficiency, the vibration methods usually cannot detect this type of fault except when it becomes severe and approaches the failure of the system.

Bearing temperature monitoring was historically thought to be insensitive to early mechanical faults (Kurfess, et al., 2006; Randall, 2011; Watson, et al., 2007) and was only used as an indication for loading and speed condition. The focus on most of bearing temperature monitoring was concentrated on the thermal equilibrium temperature that is attained after the balance between the heat generation and the heat dissipation rates. However in the current study, the fault effect on the temperature rise curve before reaching the thermal equilibrium condition will be studied and used to assess some rolling element bearing faults.

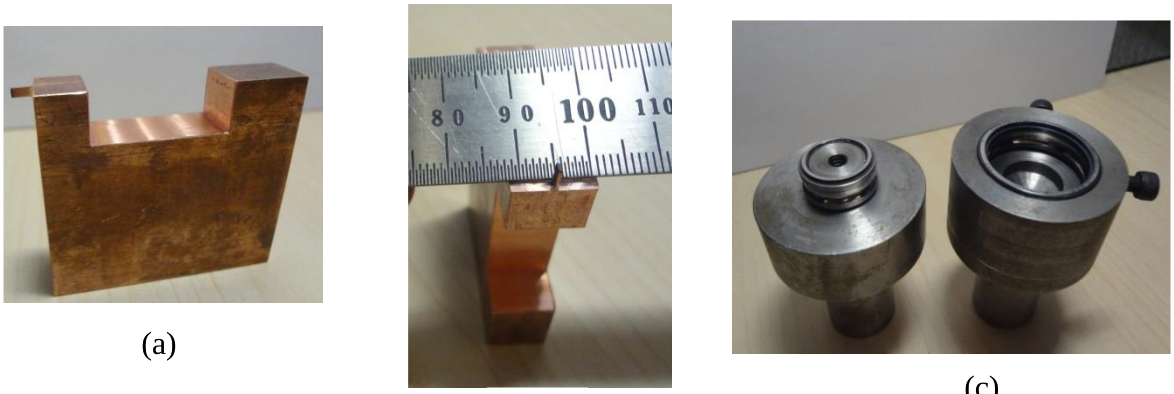
In this chapter, some of the real bearing faults will be tested using both vibration and temperature methods at different fault severity levels.

6.2 Fault Seeding

For the purpose of studying and testing bearing fault detection and health assessment methods, the two types of bearing faults, IFT and NIF faults, are produced in the radial deep groove ball bearings (Type: SKF 6202/16 RSH). The characteristics and dimensions of the bearings and the test rig were previously mentioned in Chapter 4. This section describes the fault types and fault seeding process.

6.2.1 Impulsive faults: inner and outer race spalls

Spalls of different sizes are seeded in six ball bearings using the Electrical Discharge Machining (EDM) technique. The electrode used in seeding the faults and the clamp fixtures in addition to the faults seeded in the three pairs of bearings are shown in Figures 6.1 and 6.2. The bearing fault shapes and dimensions are listed in Table 6.1 and depicted in Figure 6.3.

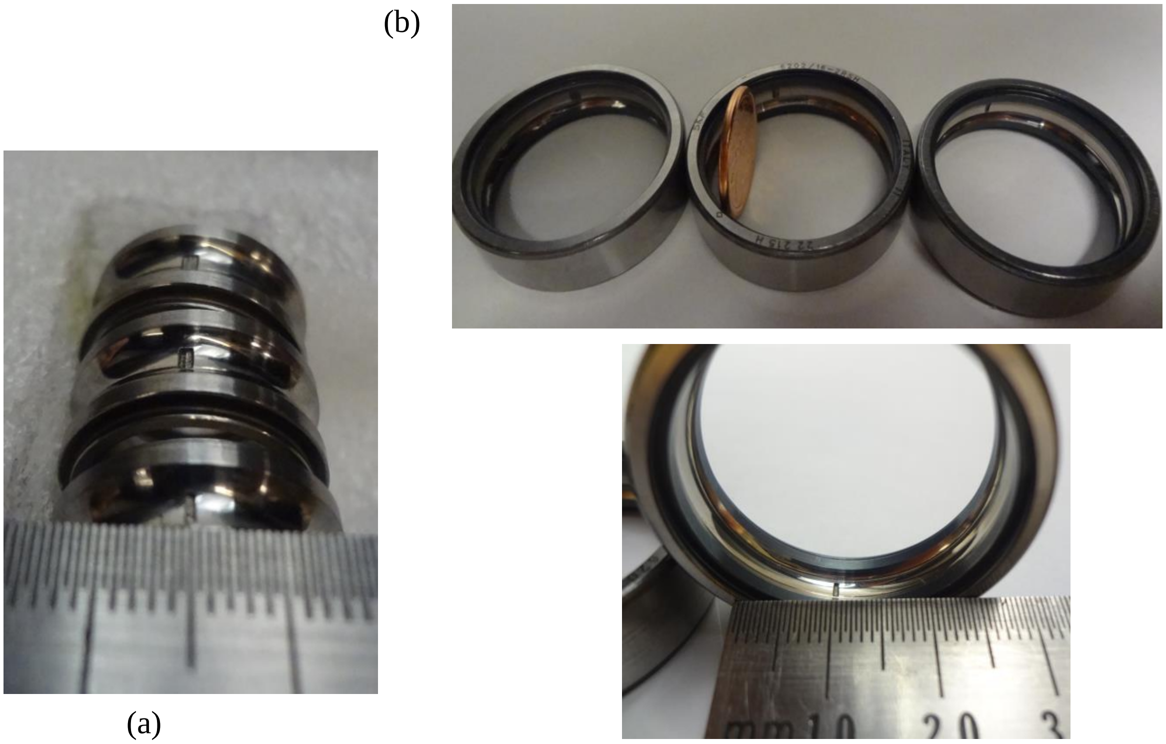


(a)

(b)

(c)

Figure 6.1 a) The EDM electrode, b) 0.6 mm electrode tip, and c) fixtures for both bearing inner and outer races



(b)

(a)

(c)

Figure 6.2 a) Three faulty inner races with spalls of 0.6, 1 and 1.4 mm wide respectively, b) three faulty outer races with spalls of 0.6, 1 and 1.4 mm wide respectively, and c) close-up view of the 1-mm-wide outer race spall

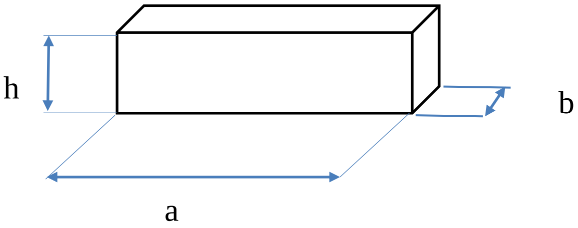


Figure 6.3 Bearing spall fault shape.

Fault category	a	b	h
Small	2.0 mm	0.6 mm	0.2 mm
Medium	2.0 mm	1.0 mm	0.2 mm
Large	2.0 mm	1.4 mm	0.2 mm

Table 6.1 Dimensions of the seeded bearing spalls.

6.2.2 Non-impulsive faults: lubricant deficiency

This type of faults is seeded by changing the bearing lubricant quantity during tests to study the effect of the amount of the lubricant on both bearing vibration and temperature signatures. Each time in the study when the lubricant quantity is changed the following procedures are repeated:

1. Empty the lubricant from the bearing completely, clean the rings and the rolling elements using cleaning alcohol, and then dry the bearing with compressed air.
2. Re-lubricate the bearing following the procedure advised by the manufacturer (SKF, 2003).
3. After re-lubrication, run the bearing at low speed for at least two hours to ensure the uniform distribution of the lubricant in the bearing internal space.

The quantity of lubricant is scaled and calculated using a standard scale and a syringe as shown in Figure 6.4.

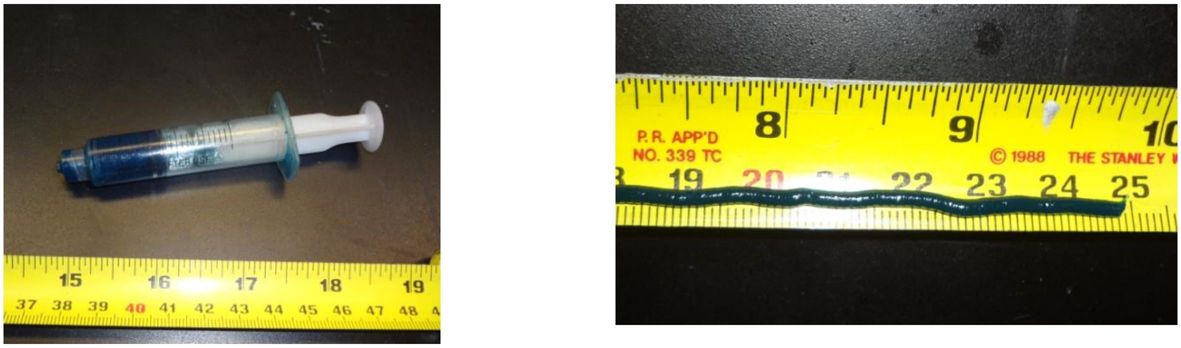


Figure 6.4 Lubricant quantity measuring tools

6.3 Temperature-Based Bearing Condition Monitoring

According to the bearing fault classification discussed in Section 6.1 the effect of each type on the bearing transient temperature will be studied independently. Despite an analytical model for normal bearing transient temperature was addressed successfully in Chapter 4, using this model while taking into account a possible bearing fault requires adding some necessary modifications. Due to the multitude and the entanglement of fault effects on bearing transient temperature three proposed simplifications are presented in the current section. The first is to assume that the bearing impulsive fault can be considered as an oscillating load that will affect both the average area of contact and the bearing total dry friction, the first module of the proposed model in Chapter 4. The difference of the average area of contact for a healthy and faulty bearing that has an IFT type can be calculated via dynamic modeling for the fault impact force in addition to a FE modeling for a single oscillating radial force acting on a ball that rolls between two races. The second simplification is to consider that the lubricant quantity linearly affects the viscous drag force and hence the total bearing hydrodynamic friction, the second module of the proposed model in Chapter 4. The last simplification is to consider that the lubricant as a heat sunk inside the bearing is uniformly distributed in the bearing internal space such that the surface of heat convection between the bearing internal parts and the lubricant is linearly decreasing with any decrease of the lubricant quantity. This last assumption will be used for evaluating the effect of possible lubricant deficiency problem on bearing heat transfer module, the third module of the model in Chapter 4.

The previous simplifications will be used to predict the necessary modification to the bearing temperature model for the two main types of bearing faults, the IFT and the NIF, in the next two sub-sections.

6.3.1 Effects of impulsive faults on bearing transient temperature

As mentioned before, faults like localized pitting, spall and initiating cracks in rolling element bearing are considered as impulsive faults (IF faults). The presence of one of these types of faults will affect the heat generation rate and the bearing temperature. The effect of the fault that has an impulsive nature can be explained by re-visiting the heat generation model and its sub-models described in Figure 3.7 as follows:

1. The bearing dynamic model – explains that the impulsive fault will change positions, speeds and accelerations of the boundary elements that establish the area of contact. This also leads to the change of stresses. It can also change the equations used by the sub-model if the resulting stresses exceed the elastic limit of the contacting bodies.
2. The viscous drag sub-model – describes that the change in the relative speeds will result in a change of the lubricant flow characteristic, i.e. new values of Reynolds's number and consequently a new drag coefficient that can be calculated by Equation (4.3). This leads to new values of the resulted viscous drag force that is calculated by Equation (3.68).
3. Lubricant film thickness sub-model -- describes that the change of loads and speeds calculated at each of contact area points results in a new lubricant film thickness.
4. Hydrodynamic shear stress sub-model – explains that the integrated shear stress due to hydrodynamic of full film lubrication regime will change due to the changes of both relative speeds and lubricant film thickness. This is concluded by Equations (3.45 a) and (3.45 b).
5. Heat transfer model – explains that the impulsive faults may cause the change of lubricant thermal properties when the heat source cannot be considered steady and the oscillating source effect should be included.

In general, it can be seen that the presence of impulsive faults will affect almost all the factors involved in bearing heat generation and heat dissipation rate calculation. However, assuming the simplifications discussed in the previous section, the effect of the impulsive fault will be calculated in two steps: the first is the calculation of instantaneous impact force using the impact dynamic equations and the second is using a FE model to calculate the change of friction stress and area of contact due to an oscillating load that has the same amplitude as the calculated impact force in the first step. These steps will be used for the calculation of the IFT type fault effect on bearing transient temperature as presented in the following two sub-sections.

6.3.1.1 Impact force, F_{imp}

A localized bearing fault is depicted in Figure 6.5 to illustrate the passing of the rolling element over the localized fault, a spall or dent fault. For a more realistic representation of the fault, Figure 6.5 depicts a 1:10 scale representation of the real dimensions of the rolling element, the contact area and the steady state deflection of an outer race of the bearing that was modeled in Chapter 4. Figure 6.5 also depicts an outer race spall fault which has a 0.6 mm width, w_f and 0.3 mm depth. The analysis results also apply to inner race faults. The major result of IFT fault is the impulse force, F_{im} that is exerted by the fault edge on the rolling element.

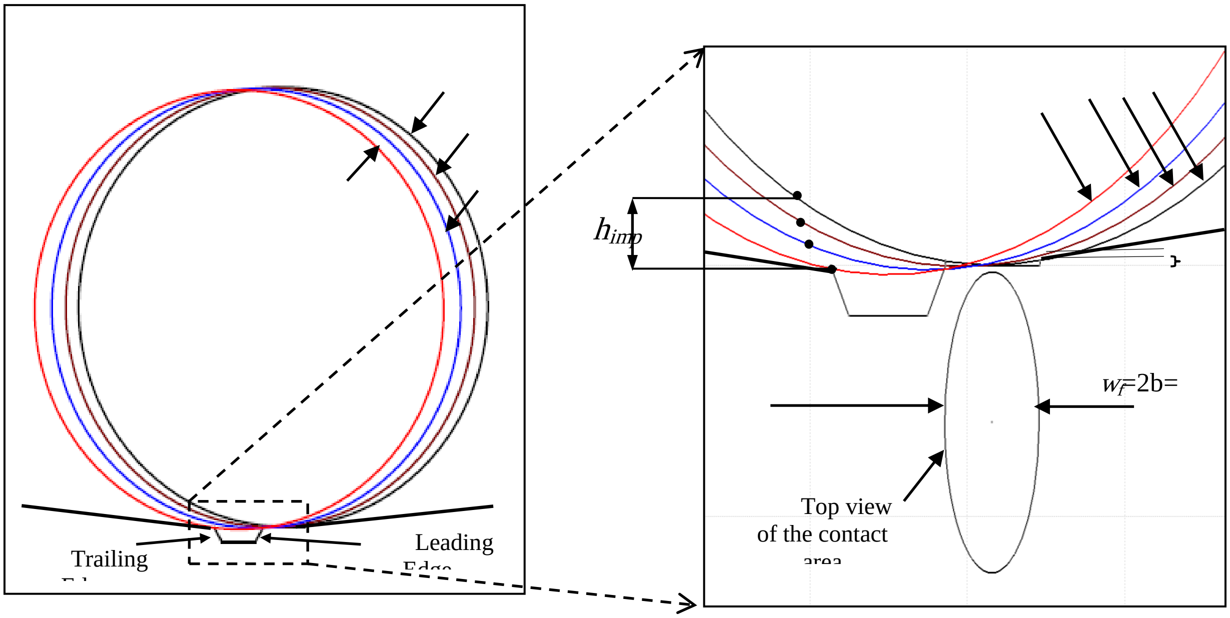


Figure 6.5 Ball impact to the fault edge (scale 10:1 of real dimensions)

The impact force resulted from passing of the rolling element over a localized fault can be solved using the simplifications suggested by Popov (Popov, 1990) described below:

1. At the instant of impact, the material of the bearing race behaves elastically, and no energy is dissipated, i.e. all the kinetic energy of the rolling element is converted into a strain energy.
2. The deflection is proportional to the load.
3. The rolling element is considered as a rigid body and the deflection only occurs in the bearing races.
4. The bearing races' inertial resistance to deflection is negligible.

With the above assumptions, the resulted impact force can be calculated immediately after the impact between the rolling element and the fault trailing edge if the bearing effective radial load at the impact point and the distance travelled by the rolling element before impact are known. Hence, the impact force is:

$$F_{imp} = Q_s \left(1 + \sqrt{1 + \frac{2h_{imp}}{\delta_0}} \right) \tag{6.1}$$

where $F_{imp}$is the resulted impact force at impact point of the contact, $Q_s$is the load applied at the same azimuth angle in the absence of a fault, $h_{imp}$is the length (height) traveled by a rolling element point before impacting the fault trailing edge and δ_0 is the deflection resulted by the passing of the rolling element on the race in steady state operation as shown in Figure 6.5. In this figure the four circles, C_1 to C_4, depict the motion of the rolling element when passing over the incipient fault, associated with four different instants starting with circle 1, followed by 2, then 3 and finally 4. Circle 1 represents the position of the rolling element when its leading edge of the contact area is just at the leading edge of the fault while circle 4 represents the instant at which the rolling element impacts the trailing edge of the fault. δ_0 is the elastic deflection caused by the steady state loading before reaching the fault.P_1through P_4 represent the four positions of the same rolling element point P on the rolling element prior to (P1 through P3) and immediately after (P4) the impact with the trailing edge of the fault. The fault size, location and the bearing geometry will be used to calculate h_{imp}.

By checking the simple geometry of the ball passing over the fault that is shown in Figure 6.5 it can be noticed that the distance traveled by the first impact point of the rolling element with the fault trailing edge, from P_1 to P_4, while P_1 is the postion of this impact point when the rolling element reaches the fault leading edge, equals approximately the fault width. Then the height of impulse in Equation (6.1) can be approximated by the width of the fault, i.e. $h_{imp} \approx w_f$. Therefore, the impulsive force at each point of the contact area between the rolling element and the trailing edge of the fault can be expressed as a function of the fault width and the static load Q_s that depends on the orbital location of the fault (Azimuth angle at which the fault exists). The next step is to use the calculated dynamic impact force as an oscillating load that is added to the bearing steady state load. The frequency of the dynamic impact load equals the characteristic frequency of the fault.

6.3.1.2 Fault impact force distribution

The impulsive force resulted from the fault impact will have an oscillating nature and can be only modeled while considering the elastic and damping characteristics of the raceway. It can be modeled as a damped driven simple harmonic oscillator that is depicted in Figure 6.6. The model considers a mass m_b that is excited by the force F_{imp} between the two raceways, inner and outer races that are modeled as two pairs of springs and dampers. The rolling element is affected by its weight ($m_b g\cos\psi$), the bearing equivalent load ($Q\cos\psi$) and the excitation impact force, F_{imp},. The fault impulsive force will have a frequency equal to the fault characteristic frequency and damped by the structure. The steady force applied on the contact area when neglecting any oscillation due to the raceway waviness is given by Equation (6.2). Then the dynamic force applied on a rolling element in the presence of an impulsive fault will be the sum of the steady state force, the body forces and the oscillating impact force as expressed in Equation

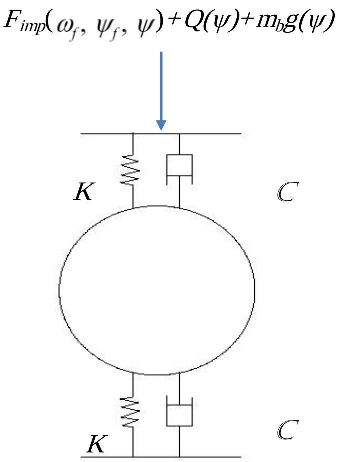


Figure 6.6 The forced non conservative oscillating system representing the motion of the rolling element between the bearing races with the fault effect.

The rolling element motion is described analytically by Equation (6.4) and can have approximate solution using several methods such as the multi-scales method or the averaging method (Nayfeh, et al., 1979; Hess, et al., 1991; Hess, et al., 1991).

$$F_{steady}(\psi)=Q(\psi)+m_b g(\psi) \tag{6.2}$$

$$F_{dynamic}(\psi)=Q(\psi)+mg(\psi)+F_{imp}(\omega_f,\psi_f,\psi) \tag{6.3}$$

$$m_b\ddot{R}(\psi)+\mathbb{C}\dot{R}(\psi)+K\delta(\psi)=Q(\psi)+m_b g(\psi) \tag{6.4}$$

In Equation (6.4), R is the position vector of the contact surface, $\mathbb{C}$ is the damping coefficient for the bearing ring material, δ is the deflection at a certain orbit angle, ψ_f is the fault orbital position angle, ω_f is the fault characteristic frequency and K is the stiffness of bearing rings' material. This equation assumes that the stiffness and damping coefficients are the same for both inner and outer rings. It shows that both the velocity vector and the deflection will change due to the orbital position of the rolling element.

Assuming that the contact between the ball and both inner and outer rings is Hertzian, then the deflection in contacting bodies and the stress distribution can be calculated depending on the radii of curvature of the bodies and the applied loads.The effect of dynamic loading in Hertzian contacts is studied by Hess and Soom (Hess, et al., 1991). Their quantitative study showed that the application of an oscillating force on a non-conservative system results in a 10 % decrease of average deflection and average area of contact that are formed by the same load in its steady state.

In order to study the bearing temperature change under the effect of an impulsive fault, the friction change due to the oscillating load should be analyzed and the change in both the average contact area and the deflection should be approximated. The finite element methods can be used to solve this problem using the ANSYS software.

For simplicity, the problem is modeled as a one sphere rolling between the two straight rails that have the same length and curvature of the bearing rings. The first step is to define the geometry and the material for the rings and the rolling element as shown in Figure 6.7. The three bodies are then meshed taking into consideration the mesh refining at the contact area zones. The load definition and the calculations will be done twice: the first time when the rolling element is under the bearing equivalent steady load and the second time after imposing the impact force that has a frequency equal to the fault characteristic frequency and a damped decaying frequency depending on the steel restitution coefficient. The contact pair is defined as frictional contact for both the inner and the outer rings contacting with the rolling element as shown in Figure 6.8.

The initial condition of the problem is set to start at zero speed at which the deflection and the contact area is equivalent to the Hertzian solution for elastic contact. Then the model is run to calculate the area and the deflection resulted in the two cases. The resulting contact areas for the steady loading and the superimposed damped oscillating load are listed in Table 6.2 and depicted in Figure 6.9.

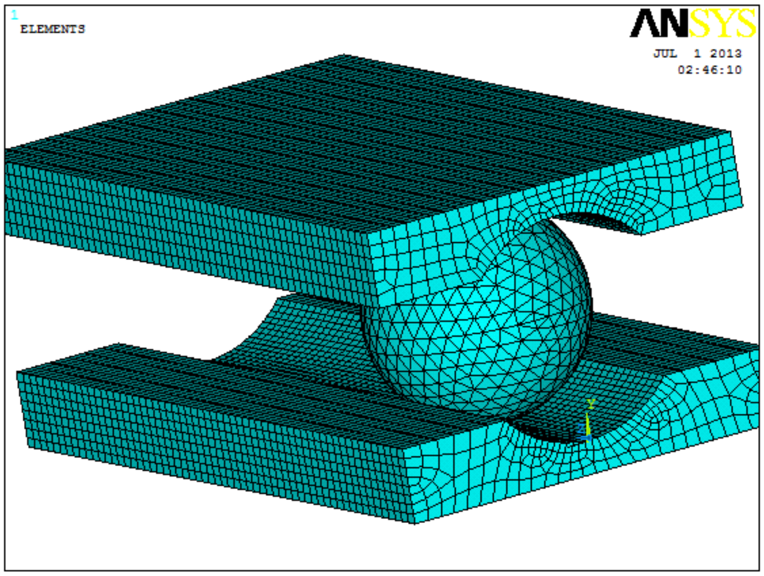


Figure 6.7 Finite element modeling of the bearing friction under decaying oscillating load using ANSYS software.

Condition	**Normal**	**Fault impact force added**
Total Area of contact, mm^2	0.162	0.149

Table 6.2 Areas of contact for steady state loading and with impact oscillating force added calculated by the ANSYS finite element software bearing friction using ANSYS software.

The finite element program also calculates the friction shear stress that can be used directly in heat generation calculation to solve for the bearing temperatures. Figure 6.9 shows the resulted friction shear stress calculation process in ANSYS.

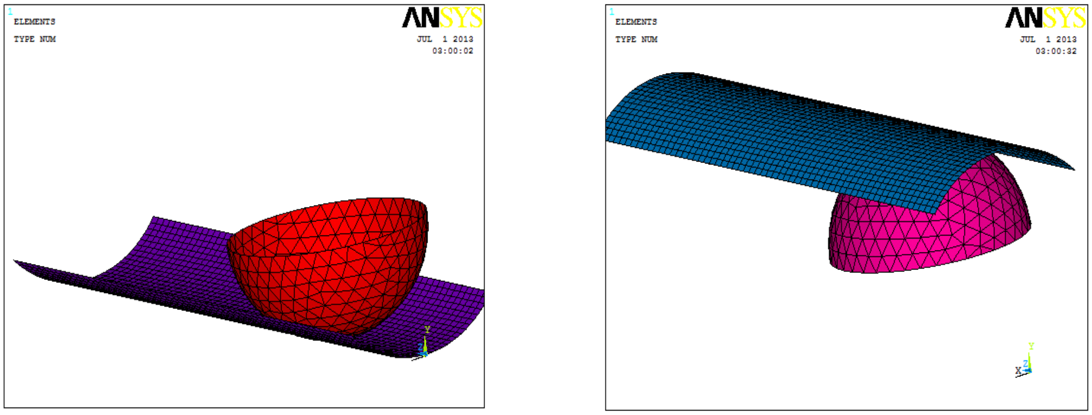


Figure 6.8 Modeling of the contact pair for both inner and outer rings with the rolling element using ANSYS software

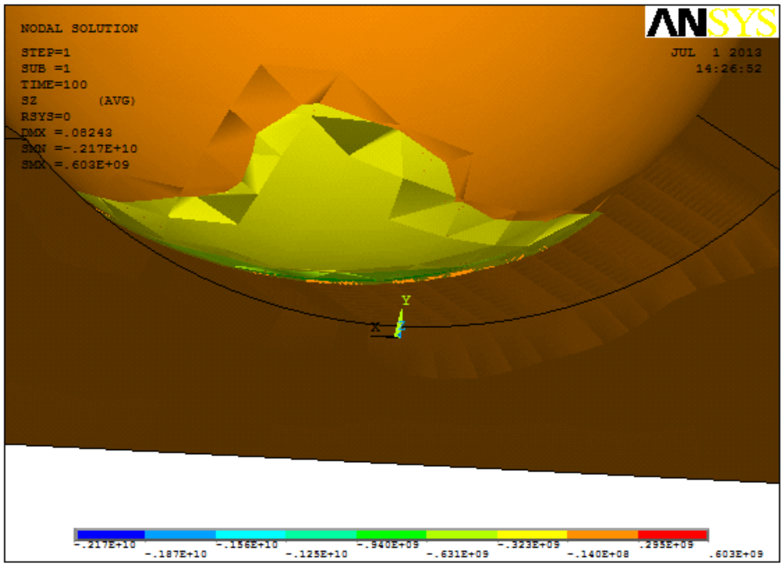


Figure 6.9 Friction shear stresses at the contact area calculated using the ANSYS FE software

6.3.2 Effects of non-impulsive fault on temperature

The lubricant related bearing faults can cause severe damage when reaching a mature level. Faults like the shortage of lubrication because of leakage and the lubricant deterioration or loosing viscosity are very popular. On contrary to the localized faults, these faults usually occur gradually and do not cause the impulsive vibration signature that can be an early sign of the fault initiation. Although bearing equilibrium temperature monitoring is good for this type of faults, reaching the value of equilibrium temperature can take more than 10 minutes and in some applications this time duration a severe damage could develop. Therefore, the use of a temperature rise curve for the bearing surface is required for early detection of the bearing NIFs.

A bearing NIF can affect two or more systems involved in temperature rise process. Firstly the reduced lubricant quantity in the bearing internal space will decrease the viscous drag friction according to Equation (3.68) and the resulted viscous drag heat according to Equation (3.72). Secondly, the heat dissipation rate will decrease due to the reduced lubricant latent heat capacity. The net change of the bearing temperature rise due to the fault will be the consequence of these two changes which depend on the amount of lubricant in the bearing.

Although both the lubricant shortage and the lubricant properties deterioration can lead to the increase of heat generation rates due to the decrease of lubricant film thickness, It has been shown in several studies (Zhu, et al., 2011; SKF, 2003) that the quantity of lubricant that is required to form the lubricating film is very small. Then, for the case of bearing incipient faults the major effect on bearing temperature is more probably because of two affecters or both of them. The first is the decrease of viscous drag friction as a result of lubricant quantity decrement while the second is the decrease of heat transferred out of the system through lubricant. Both of the two affecters are regarded in the bearing temperature model. The change of the viscous drag friction is calculated by the second module and the lubricant quantity effect on heat transfer rates is regarded in the third module of the bearing transient temperature model.

6.4 Experimental Results

The experiments including the monitoring of bearings with IF and NIF using both temperature and vibration monitoring methods. One of the purposes of using both the temperature and vibration methods is to examine their effectiveness in detecting and assessing incipient and mature faults. Both short time and extended tests will be accomplished for temperature monitoring to show the effect of the fault on both the thermally transient and thermally equilibrium conditions. The short period test is only for five minutes while the long test could last for 50 minutes. The following procedure will be followed in all tests:

1. The temperature measurement accomplished in all tests will be compared with the theoretical results obtained by the model proposed in Chapter 4 and with the measurement of a healthy bearing mounted in the same place and running under the same conditions at the same ambient temperature.
2. For the non-impulsive faults, the theoretical results will be modified according to the lubricant quantity change.
3. For all the tests that will be shown in this section, the vibration signal is acquired by two accelerometers with frequency measurement limits up to 60 KHz and 20 KHz, respectively. The higher frequency accelerometer is fixed at a vertical angle with respect to the shaft axial direction while the other one is fixed in a horizontal angle. The two accelerometers are fixed to the bearing housing using thread fixing type. The magnet fixation is avoided as it is affected by system vibration especially at high speeds.
4. The temperature monitoring is done using both a FLIR-E40 thermal camera and an OMEGA-OS211 infrared sensor described in detail in Chapter 4.
5. For each test, the log history of the operation hours and speeds are recorded as this data is important for impulsive fault development.
6. When completing a test for the first time for a bearing or after re-lubricating process, the bearing is run firstly for at least two hours at speed 30 Hz, i.e., 216,000 revolutions to

ensure uniform distribution of the grease lubricant in the rolling contact (SKF, 2003; Johnson, 2009). Then the bearing is let to return to the ambient temperature before running the test.

7. The temperature rise curves for different bearing faults will be compared to that calculated using the temperature model described in Chapter 4 and to the curve measured for the healthy condition under the same load and speed condition.

6.4.1 Tests of bearings with impulsive faults

Two bearings with different sizes of outer race fault are examined by both vibration and temperature monitoring methods. The vibration data acquisition system and the temperature monitoring system described in Section 4.3 will be used. The two bearings have spalls of the same depth (0.2 mm) but different widths, 0.6 mm and 1.4 mm, respectively as shown in Figure 5.2. The following three groups of tests will be conducted:

1. Short time tests.
2. Extended tests.
3. Matured fault test.

The duration of the short time tests are fixed to 300 seconds. The durations of the other two groups are situation-dependent to be detailed later. For the vibration method, as both the short time and extended tests give the same vibration signal, only a short time sample will be considered. The results of short time tests using vibration methods have been presented previously in Chapter 5. Hence, the vibration methods will only be used in current section for checking the fault maturity. The onset of fault maturity is detected by continuously monitoring the envelope spectrum at the fault characteristic frequency. When a spall fault exists in the raceway such as the fault depicted in Figure 6.5, the edges of the fault are considered as stress concentration zones. Several studies such as (Branch, et al., 2009; Randall, et al., 2011; Randall,

2011) have shown that, even with light loading operation of the rolling element bearing, the stress concentration can cause the exceeding of elastic limit. In other words the fault edges will deform plastically according to the resulted very high stress in the stress concentration zones. Regarding Equation (6.1), the magnitude of the impact force depends on the fault width that decreases gradually due to the plastic deformation of the area around the fault edges.In this study, a fault is defined as mature if the average amplitude of the envelope spectrum calculated over five seconds decreases to 40% of its reference value measured at the first bearing test. This definition of maturity may not be generalized because the time required to reach the fault maturity depends on operating conditions..

For the temperature monitoring methods, the short time test is for measuring the temperature rise curve of the bearing inner race side surface during the first period of operation. In this test the temperature will initially rise quickly as the heat generation rate is much higher than the heat dissipation rate. The rate of ascent of the bearing temperature decreases gradually with time till the system reaches a thermal equilibrium condition. However, reaching the thermal equilibrium condition is always very slow and will not be reached for the current system before 45 minutes of test operation at a 100 Hz shaft speed. The purpose of the extended tests is to reveal the effect of the fault on the bearing thermal equilibrium temperature.

6.4.1.1 Short time test

In this test, the transient temperature rise curves are acquired by the thermal camera and the infrared sensor for 5 minutes. The first test for an incipient stage fault was only run for the initial run-in time for grease uniform distribution.

The snapshots taken by the thermal camera for healthy, outer race faulty and inner race faulty bearings at 5, 10, 50 and 100 seconds from the beginning of the test are shown in Figure 6.10. The temperature rise curves for the inner race side surface of bearings with respect spalls of 0.6 mm and 1.4 mm in width are shown in Figure 6.11. The figure also shows the temperature rise curves measured for a healthy bearing at the same operating conditions and the temperature rise curve calculated by the bearing temperature model.

Figure 6.12 shows the temperature rise curves for a bearing with an inner race fault and a bearing with an outer race. The inner race and outer race fault sizes are identical. The temperature rise curves for the healthy bearing at the same operating conditions and the calculated temperature rise curve using the analytical model are also presented in the figure..

It can be observed from Figures 6.11 and 6.12 that temperature rise curve for the healthy bearing increases faster than the curves of faulty bearings. The difference between the transient temperature for bearings at different condition starts with high value then it decreases with time. It will be shown later that bearing temperature will reach the same value after reaching the equilibrium condition.

It can be noticed also from Figure 6.11 that the deviation of the temperature rise curve is higher for the larger sized fault. Also, Figure 6.12 shows that the temperature rise curve for inner race fault has a larger difference from the reference curve of the healthy bearing when the two different faults have the same size (1 mm in this case).

To better examine the effect of an NIF on the temperature transient behavior, the temperature rise curves related to the inner and outer race faults are subtracted from the temperature rise curve of the healthy reference bearing. The results are plotted in Figure 6.13 (a) and 6.13 (b) respectively. The following can be observed from the figure:

1. The rapid upward transient trends of the first 30 seconds provide sufficient information to discern the bearing conditions. This makes it possible to extend the detection horizon for early fault detection.
2. Temperature readings for inner and outer race faults are lower than that of healthy bearing. However, the difference between the healthy bearing temperature and these two faulty bearings' temperatures decreases with time. The decay rate for this difference is higher for the outer race faulty bearing than for inner race faulty bearing. Noting that the fault size for outer race is bigger than that for the inner race, the difference in decay rate could be used to devise a fault size or a fault severity indicator.

The figure also shows that the temperature difference in the thermal transient condition starts at high value and reaches a maximum of 6.5 degrees for the inner race faulty bearing and 5.5 degrees for the outer race faulty bearing. After the first 50 seconds of the test, the temperature difference decreases gradually.

Though the temperature of the bearing with outer race fault is approaching the reference temperature of the healthy bearing faster than the temperature of the bearing with inner race fault, it will be shown in next test that the temperature for the three bearing conditions will reach almost the same equilibrium temperature.

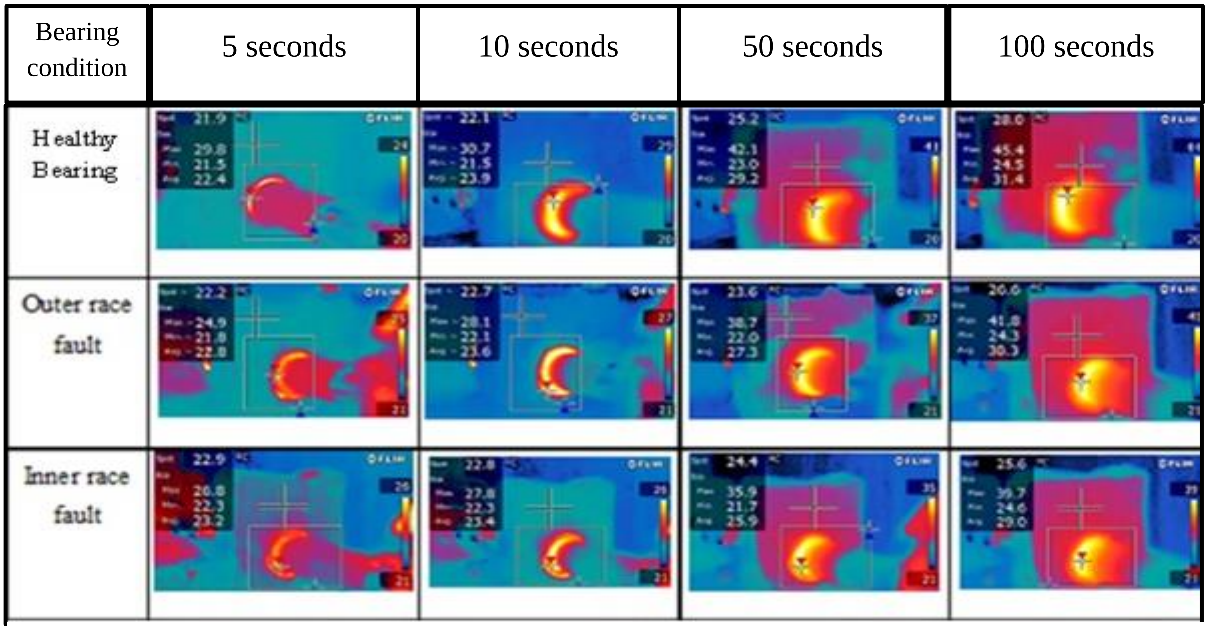


Figure 6.10 Snapshots taken by a FLIR-E40 thermal camera for the side surface of the inner race of three bearings at three different conditions, healthy, faulty outer race and faulty inner race, recorded 5, 10, 50 and 100 seconds into the tests running at 100 Hz shaft speed.

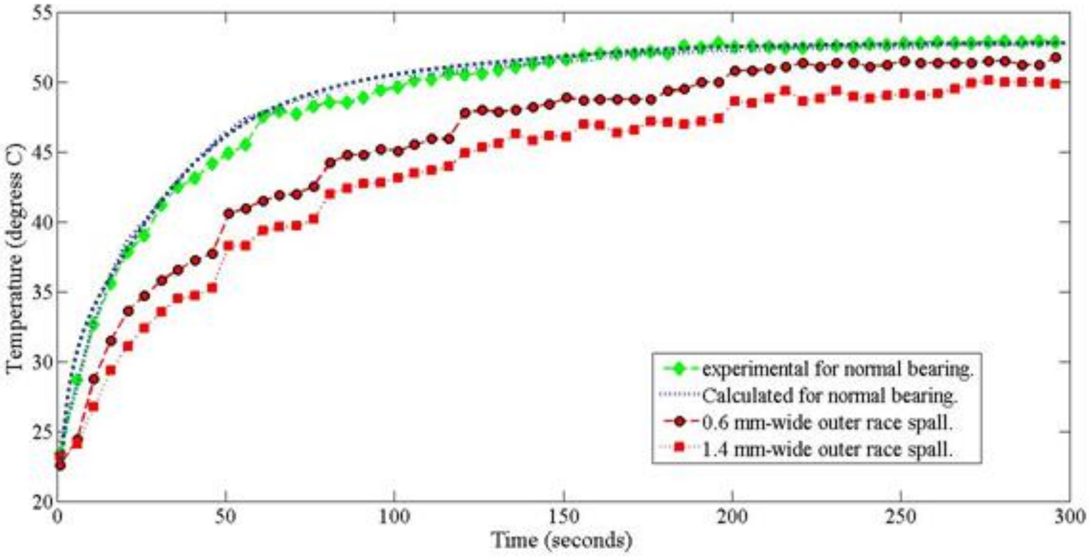


Figure 6.11 The temperature rise curves measured for 300 seconds for the healthy bearing, two bearings with different sized spalls, and the calculated curve for the healthy bearing.

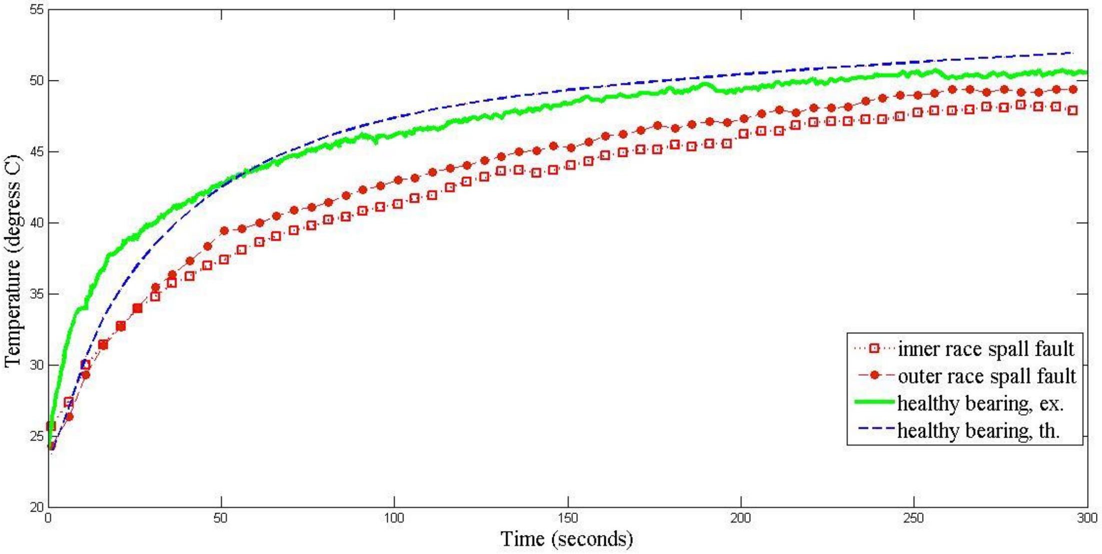


Figure 6.12 The temperature rise curves measured for 300 seconds for the healthy bearing, two faulty bearings (one with an inner race spall and the other with an outer race spall of the same size, 1 mm in depth), and the calculated curve for the healthy bearing.

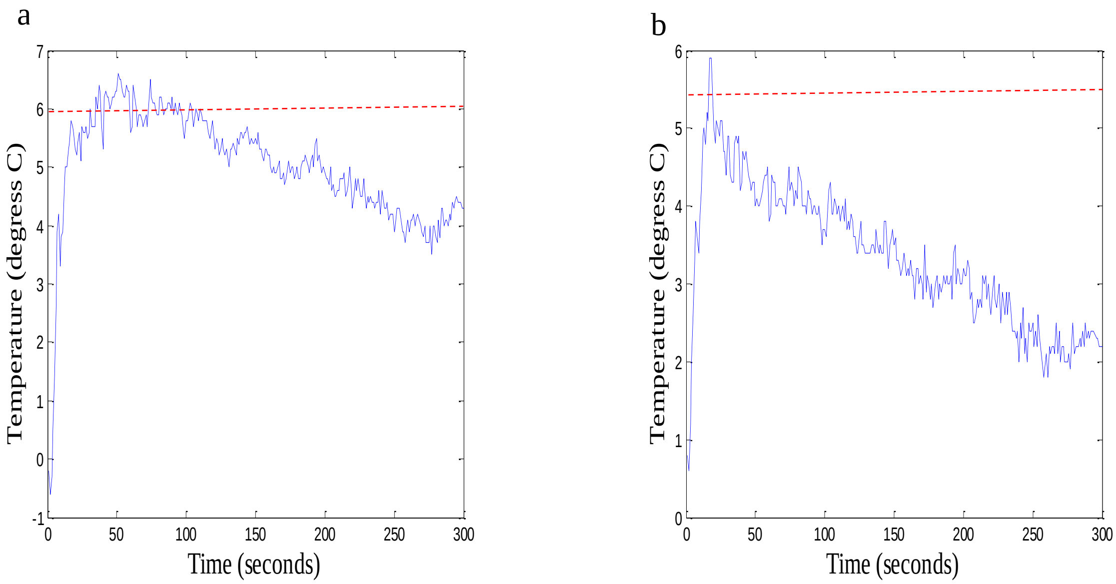


Figure 6.13 Temperature rise differencesbetween: (a) the temperatures of the reference healthy bearing and the bearing with an inner race fault, and (b) thetemperatures of the reference healthy bearing and the bearing with an outer race fault.

6.4.1.2 Extended tests

It generally takes a long time for the bearing system to reach equilibrium condition. This time duration depends on the heat generation and heat dissipation rates and can vary even due to minor changes in thermal system surrounding the heat source, i.e., the bearing. The extended tests for the temperature of the test bearing external surfaces are also carried out for a healthy bearing and faulty bearings with inner and outer race spalls.

The results show that in spite of the differences in temperature rise curves for the healthy and faulty bearings, the equilibrium temperatures attained at the extended tests are almost the same for the different bearing conditions. Two possible reasons for the same equilibrium temperatures are: a) the change in equilibrium temperature caused by a small fault like these used in the current study is too small to make a big difference in the system equilibrium temperature, and b) the accuracy of the temperature measuring devices may not be sufficient to reveal the difference.

Most of temperature measuring devices have at least ± 1° C error which can be considered a big error with respect to the change caused by incipient faults.

Figure 6.14 shows the temperature rise curves for both a healthy bearing and a faulty bearing with a 1 mm wide spall in the inner race. The Figure depicts a clear divergence in the temperature rise curves of the healthy and the inner race faulty bearing in the first 100 seconds. After this the discrimination between the two curves will start to be difficult till it becomes impossible to distinguish this difference, almost after 500 seconds from the beginning of the test. And so it can be reasoned that the transient thermal behavior of the bearing surface can be exploited in bearing fault detection. If the system has already progressed to the equilibrium condition before the fault occurrence, it will be impossible to detect an incipient fault. This will fix the usage of temperature based methods to the assessment of bearing only at its starting stage of operation. To avoid this limitation a new technique for stimulating the thermal transient behavior will be studied and applied for bearing fault detection in Chapter 7.

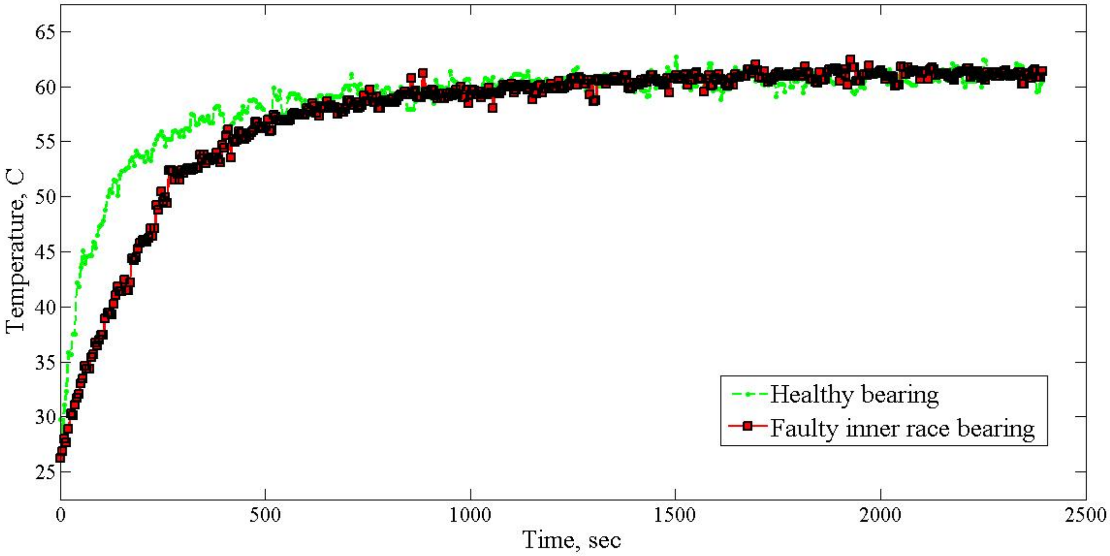


Figure 6.14 Extended test results of healthy and faulty inner race bearings.

6.4.1.3 Matured fault test

A localized bearing fault passes through several stages before a complete failure occurs. Supposing that a spall fault is caused by the debris of a small part of the bearing race due to either manufacturing defect or operation problems, the spall fault will start at some size depends on the cause of the fault. During the operation of the bearing, the rolling element will continuously hit the edges of the fault. Even in the lightly loaded bearings such as the case of our tests, the impact of the rolling element with the fault edges will cause plastic deformations of these edges. Figure 6.15 shows the outer race fault at different maturity level after 100,000 and 1 million shaft revelations. Despite of the miniature size of the plastic deformation caused by the rolling element by each impact, these deformations will cause the change of the fault shape and make it blunt. This change in the fault shape will decrease the strength of the impact resulted from passing of the rolling element over the fault and will make detection using vibration methods more difficult. However, the change of the fault shape will not have the same effect on the temperature method. The difference is because the vibration methods depend mainly on the strength of the impact produced by the localized fault shape while the temperature change due to the presence of the localized fault is contributed by many sources. One of these sources is the reduction of the contact area caused by rolling element oscillation. Although this source will also be affected when the vibration impulses decrease due to the fault shape change, other sources affecting the heat generation and heat dissipation mechanisms will not be affected. Both the change of the viscous drag due to the bearing internal volume variation and the change in hydrodynamic shear stress due to the change of the pressure applied on the lubricant will affect the resulted tribology system and will result in a difference in temperature rise curve.

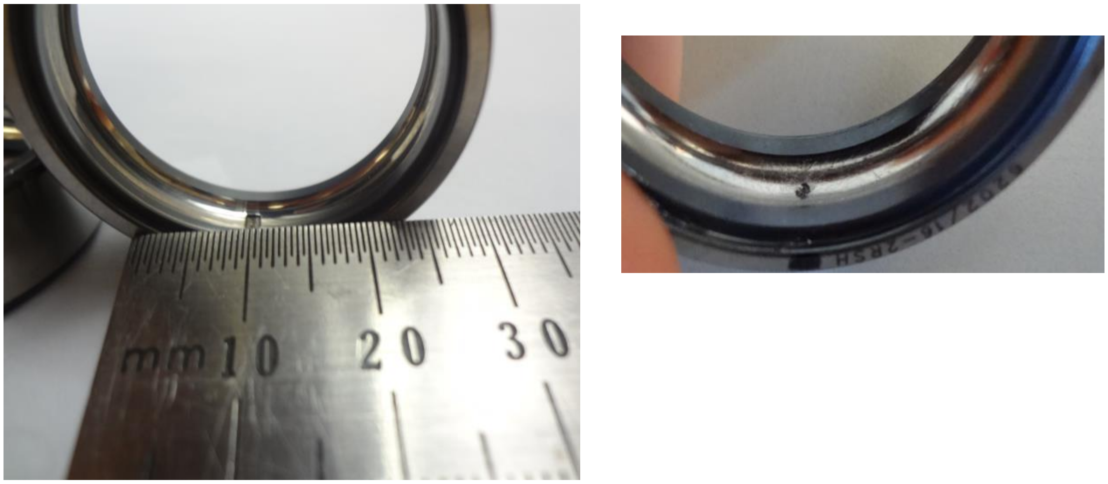


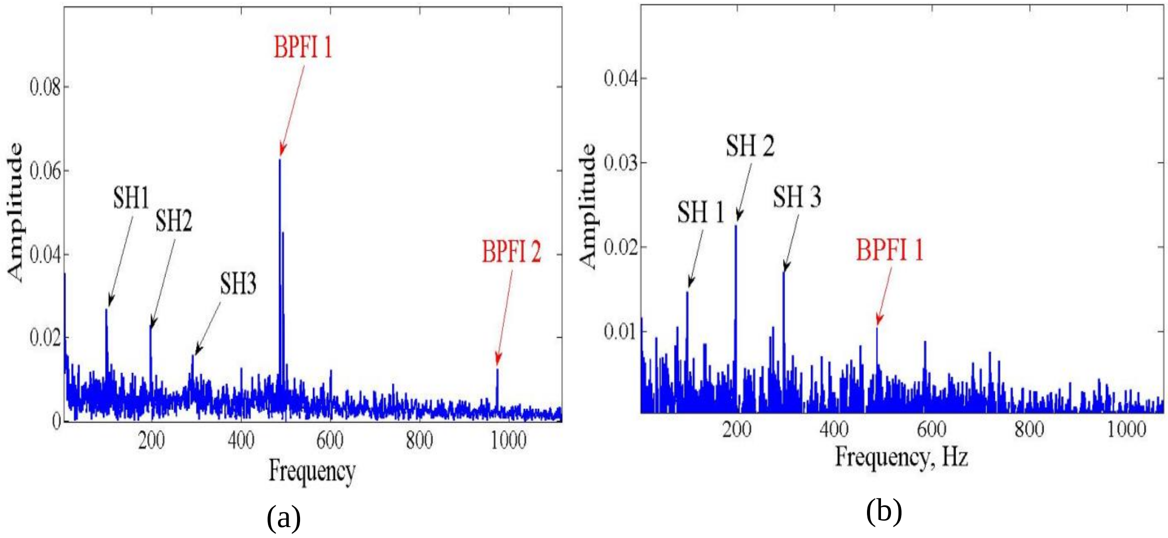
Figure 6.15 bearing outer race fault at different maturity levels: (a) after 10^5 shaft cycles, (b) after 10^6 shaft cycles

To examine the above idea, an outer race fault with initial dimensions of 1.4 mm in width, 2 mm in length and 0.2 mm in depth shown in Figure 6.2 is tested. The bearing with this fault is operated and monitored using vibration analysis method described in Chapter 5. The envelope spectrum is obtained after each 2 hours of operation. It was found that the envelope spectrum at the fault characteristic frequency started to decrease after only 3 hours of operation. The condition of maturity was reached after almost 18 hours of operation.

The envelope spectra for the bearing with different fault maturity level are shown in Fig 6.16(a) and 6.16 (b). It is noticed from Figure 6.16 that the amplitude of the envelope spectrum at the fault frequency decreases to less than 25% of its average value measured when the fault was still immature. The problem of this phenomenon is that it gives a wrong indication about the bearing health.

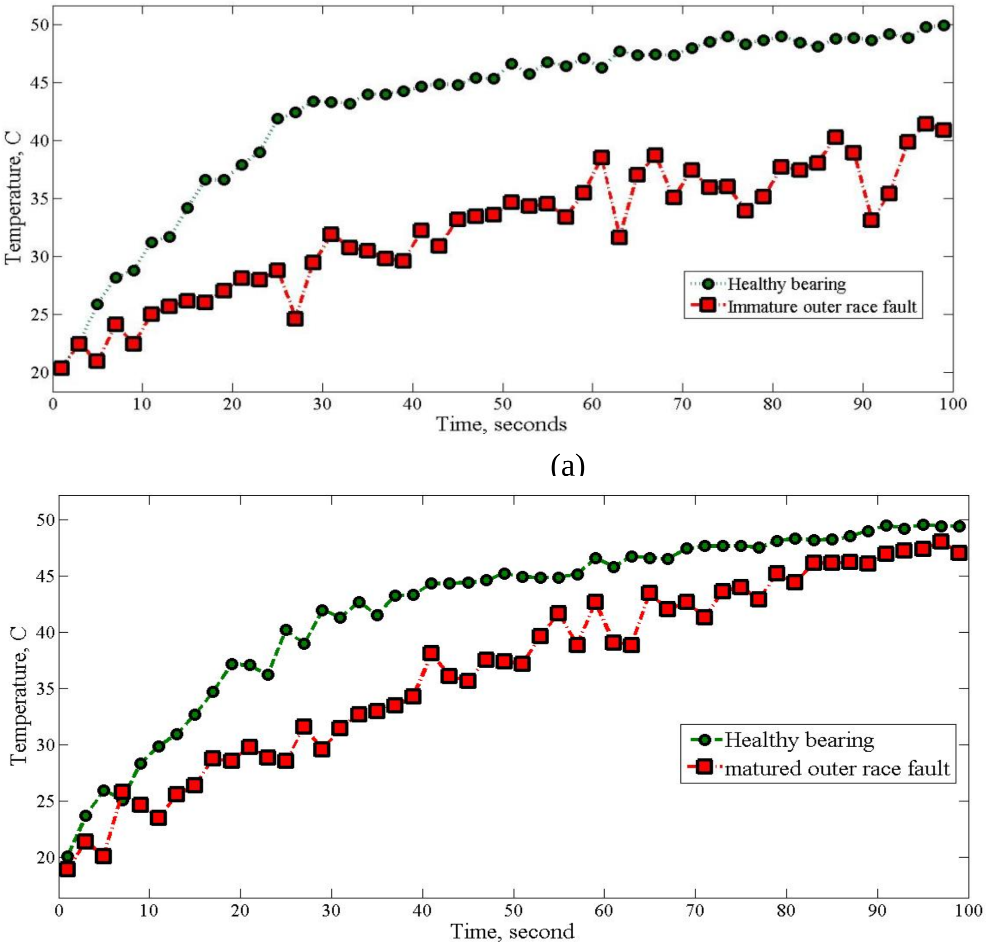
The temperature short time test method is used for the two maturity level faults with their envelope spectrums shown in Figure 6.16. On contrary to the difficulty faced to detect the mature fault using the vibration method, the temperature method showed almost the same difference between the temperature rise curves of the two maturity levels of the faulty bearing.

Figure 6.17 shows the temperature rise curves for the same bearing at two different maturity levels compared to the temperature rise curve measured for the same bearing at healthy condition. It can be noticed from Figure 6.17 that the deviation between the healthy and faulty bearings is larger in the case of the immature fault than the mature one. This deviation increases from the commencement of the test, and then makes its maximum after 34 second for the immature fault case and after 36 seconds in the mature fault case. Besides, the biggest difference between the two curves is 8 degrees in case of immature fault while it is only 5 degrees in the case of the mature fault. Despite the reduction of temperature curves deviation because of the increment of the fault maturity level, the deviations for both events are still noticeable and can be used as a mean for bearing health assessment.



(a) (b)

Figure 6.16 Envelope spectra at fault characteristic frequency after (a) 100000 shaft cycles, and (b) 6,400,0000 shaft cycles



(a)

(b)

Figure 6.17 Temperature rise curves for a healthy bearing as a reference and a bearing with outer race faults with different maturity levels (a) after 100000 shaft cycles, and (b) after 6,400,0000 shaft cycles.

6.4.2 Tests of bearings with a non-impulsive fault

The non-impulsive bearing faults do not have a strong vibration signature and will be difficult to detect during its incipient stage. The examples of the NIF include the lubricant deficiency and gradual wear. For some applications such as the high speed lightly loaded machines, the detection of the fault at its incipient stage is essential to protect the machine.

In this section, the lubricant deficiency detection at the incipient stage will be attempted using both vibration and temperature methods discussed in the previous chapters. One bearing which is free of the IF faults will be used while the lubricant quantity will be changed for each test. Again the experimental procedures and rules mentioned in the beginning of Section 6.4 will be followed and the lubricant quantity will be controlled by the devices shown in Figure 6.4. Six different bearing conditions are defined for the testing of the NIFs as shown in Table 5.4 below. The bearing conditions are chosen such that the difference in lubricant quantity from test to another is small, only from 62 ~124 mg. The case of dry bearing, case #1, and the original standard lubricant quantity, case # 6 calculated according to (Johnson, 2009; SKF, 2003), are also tested.

The results of the temperature method are displayed in Figure 6.18 showing that every lubricating condition has a different temperature rise curve that can be considered as a lubricant effect signature for the bearing. The temperature rise curves show that, for the current operating conditions and the light load, the effects of the viscous drag and hydrodynamic shear are dominant. The lack of lubricant is extreme in the case of the dry bearing. In this case the heat generation will be caused only by the dry friction. By adding a certain amount of lubricant, the dry friction decreases but at the same time the heat generated due to the hydrodynamic shear and the viscous drag will increase the total generated heat. This is not the case if the applied load is high as it can cause a dominant dry friction and the resulted generated heat will decrease by adding lubricant.

Case #	**Lubricant mass**	**Case #**	**Lubricant mass**
1	0 (Dry contact)	4	372 mg
2	62 mg	5	496 mg
3	186 mg	6	620 mg

Table 6.3 Different lubricant levels used in the temperature and vibration tests

The temperature rise curves for the different lubrication condition shown in Table 6.3 and depicted in Figure 6.18 are approaching the same equilibrium temperature. This explains why the equilibrium temperature is not a good mean for bearing early fault detection.

The results of the vibration methods are shown in Figures 6.19 and 6.20. The power spectrum using the FFT method (Figure 6.19) shows that the level of the vibration power at the bearing resonance frequency, around 8000 Hz, is high for the dry and low lubricant quantities. The envelope spectrums predicted by the ATSA method show also that the envelope spectra around the shaft frequency harmonics 100, 200, 300 and 400 are of high amplitude. The amplitudes of these envelope spectrums decrease with the increase of the lubricant.

Although either the vibration method or the temperature method gives a strong indication of fault related to the lubricant deficiency, especially for the cases in which the quantity difference is small, the simultaneous application of the two methods should make the detection more conclusive.

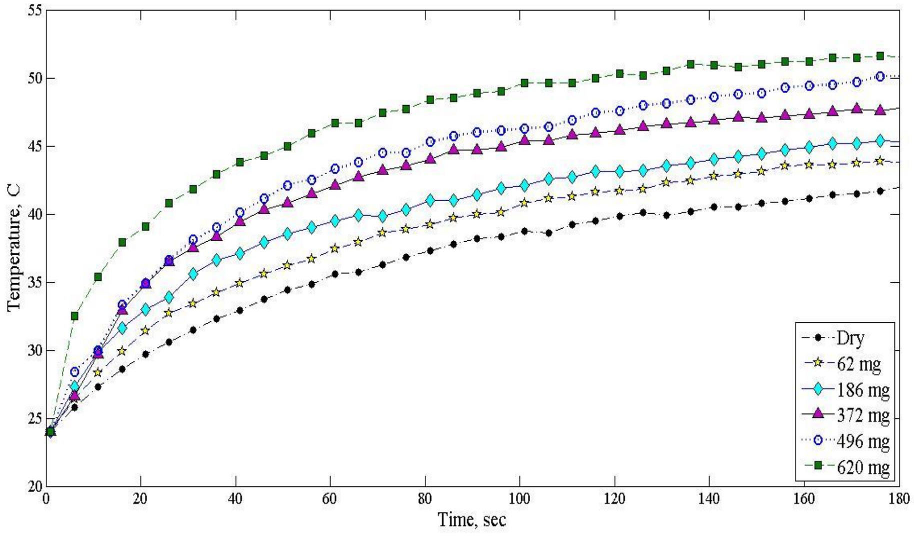


Figure 6.18 Temperature rise curves for different lubricant quantities in the same bearing and running at the same operating conditions.

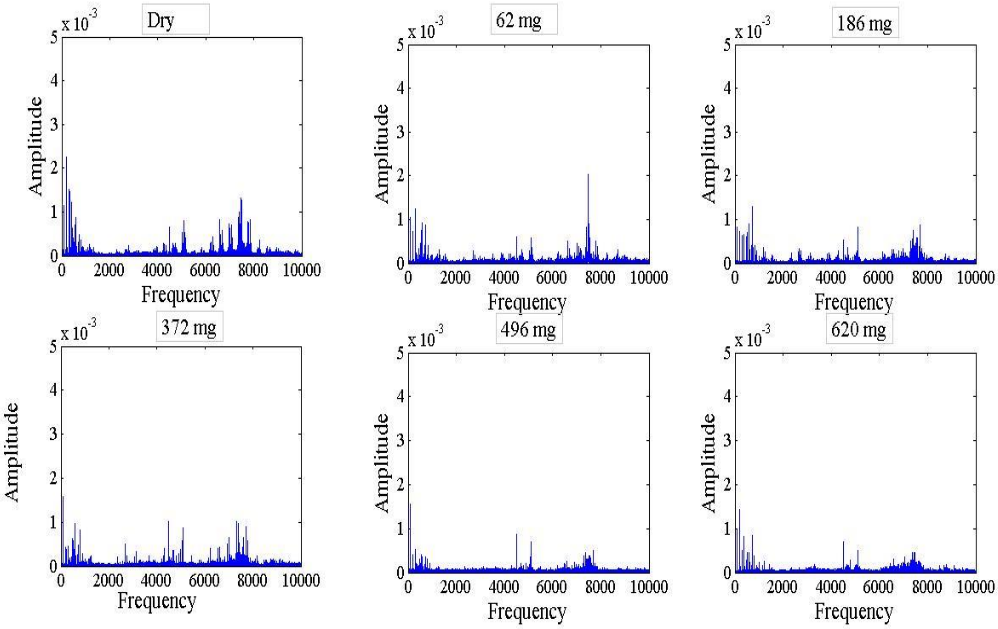


Figure 6.19 Vibration spectrums using FFT method for the bearing at six different lubricating conditions, different lubricant quantities.

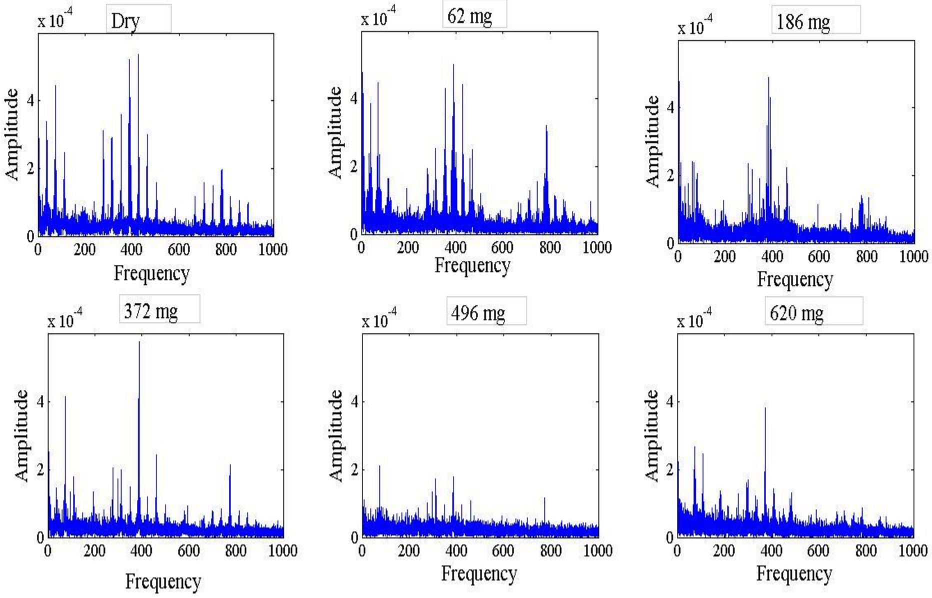


Figure 6.20 Vibration envelope spectrums using the ATSA method in 5.2 for the bearing at six different lubricating conditions, different lubricant quantities.

6.4.3 Summary of experimental results

In the current section the two types of bearing faults are examined using both temperature and vibration methods. The results of the current section can be summarized in the following:

1. A short test for the IFT and NIF types showed that the transient temperature rise curve measured at the bearing initial operation phase is a good mean for bearing health assessment. On the reverse, the measurement of bearing temperature for a long time test showed about the same temperature of equilibrium for early stage faults of both of the IFT and NIF types.
2. The vibration method that was proposed in Chapter 5 showed a very good detection capability for the incipient faults of the IFT type. Nevertheless, for the NIF type the vibration method was not effective in bearing health assessment.
3. The progress of a bearing fault of the IFT type was studied by the two methods, vibration and temperature. The results indicated that the bluntness of the spall fault due to maturity decreases the efficiency of the vibration method to discover the fault while it was yet possible for the temperature method to find it.

Chapter 7

Stimulation of Bearing Thermal Transient Behavior

The study of temperature monitoring as a method for bearing incipient fault detection in previous chapters showed that the equilibrium temperature is not sensitive to incipient fault bearing condition. However the monitoring of the initial thermally transient bearing condition showed a clear difference between the temperature rise curves of incipient faulty bearings with different fault types. This difference can be helpful to detect faults of the bearing at the initial period of operation assuming that the fault is already existing due to manufacturing or installation problems. However, if the fault is caused during the bearing operation and after the bearing has already reached the thermal equilibrium condition, the temperature-based method will not be able to detect the fault.

This chapter focuses on the stimulation of the thermal transient condition to make use of the information in the temperature rise curve signature for bearing monitoring. The idea is similar to that of active thermography (Maldague, 2001). However, there is a main difference between the active thermography method and the proposed simulation technique. In the former case, the tested item is directly heated to excite the radiation energy from its surface that will give an idea about many parameters such as the roughness and the cracks. In the proposed method, the heating source will be used to raise the temperature of the bearing housing to break the balance between the heat generation and the heat dissipation rates and hence generate a new local transient condition. This new local transient condition should have the same signature of the initial thermal transient period. When the operating temperature for the stimulated bearing is not far from the ambient temperature (the difference is in the range of 10 to 20°C), the original and the stimulated temperature rise curves can be almost identical. At higher operating temperatures the stimulated transient curve could be slower due to the thermal properties changes at different temperatures.

In order to have a transient curve that represents the bearing heat generation signature and not the heat transferred from the heat source to the bearing, the heat transfer system is modeled and solved using both the lumped mass method and the ANSYS finite element method. The purpose of the model is to ensure that the heat transferred by the heat source will only raise the temperature of the lower bearing housing, and there will be no heat flow from the heat source to the bearing.

7.1 Transient Stimulation System Modeling

The transient stimulation system, TSS, will consist of a heat source, a power supply and energizing timer. The purpose of the design step is to find out the correct specification of the heat source and the energizing time that are required for stimulating the transient behavior of the bearing system without transferring the heat to the bearing directly. The design for TSS system is accomplished by two methods, the lumped mass system that was used before as a part of the bearing temperature modeling in chapter 4 and a finite element solution using the ANSYS software.

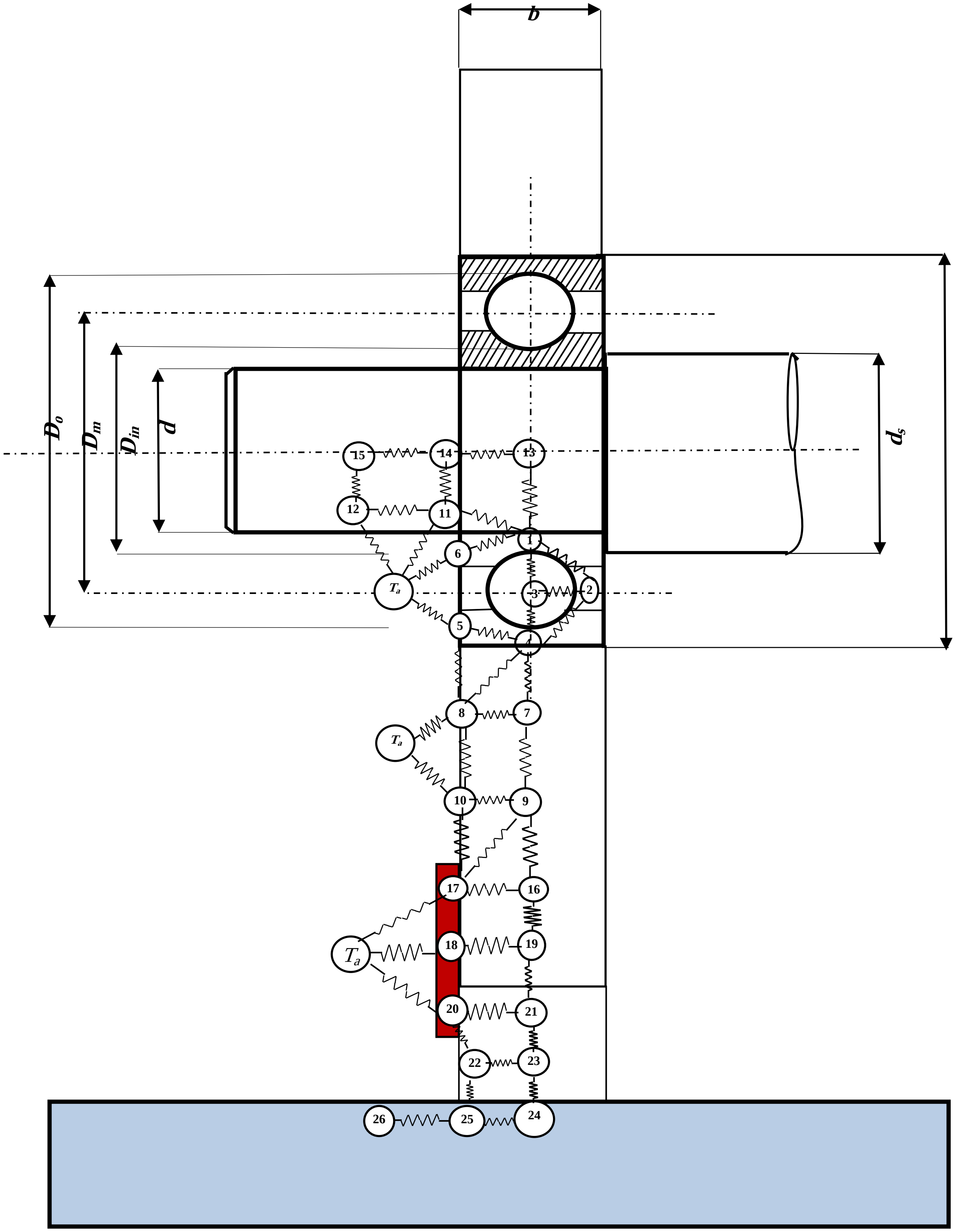


Figure 7.1 The thermal network for the TSS.

Figure 7.1 shows the thermal network to model the TSS. In addition to the points that were originally defined at the bearing temperature model in chapter 4 another 15 thermal nodes are defined. Nodes 17, 18 and 20 represent the heat source element while nodes 16-22 represent the left side of the bearing housing. Nodes 24, 25 and 26 represent the masses of the test rig lower base that will be affected by the heat source. All points that are farther than the last three points will be considered to be at the ambient temperature and are not affected by the bearing operation during the test.

The required condition for the TSS is to have rapid increase of the bearing lower housing temperature using a heat source that is energized only for a short time such that the bearing surface temperature increases due to the loss of heat balance. This condition will occur when the heat flow rate from the bearing outer race, node 4 to the lower housing, node 7 and node 8 decreases due to the increase of the temperature of the lower housing nodes. The power for the heat source and the energizing time are calculated such that it will cause the rise of the lower housing temperature and will not dominate the bearing heat generation phenomenon that is measured using thermal camera and Infrared sensor. The condition for the heat source at any moment during the heater energizing time is given below.

$$\left\{\begin{array}{l} T_{17} > T_9, T_{17} > T_{10}, T_{17} > T_{16} \\ T_9 > T_7 \\ T_{10} > T_8 \\ T_4 > T_7, T_4 > T_8 \\ T_5 > T_8 \end{array}\right\} \forall \ t \le T_{HE} \qquad (7.1)$$

whereT_{HE} is the heater energizing time that will be considered as three minutes. The thermal resistances Equations (4.13) to (4.15), predefined in chapter 4 will be used to calculate the temperature transient curve for the bearing surface when using several choices of the heater power. The heater with lowest power that could achieve the above condition in Equation (7.1) is chosen. The energizing time of the heater will be checked to ensure that is always less than the time at which $T_7 > T_4$. The results of the lumped mass model led to the choice of a 6×8 cm, 120 watt surface heater.

Figure 7.2 shows the test rig with the surface heater and the measurement devices, the thermal camera and the Infrared sensor. The resulted temperature curves using the lumped mass model for point No. 7 and point No. 4 as defined in Figure 7.1 are shown in Figure 7.3. These analytical results show that the use of two 8×10 cm, 120 watt surface heater for a 3 minutes period changes the condition of the system from the thermal equilibrium to the transient condition.

The results in Figure 7.3 indicate that the temperature at both points increase with different rates. Although the rate of temperature increase for point No. 7 is steeper, the correct energizing time will stop the heat addition from the heater before a heat flow from point 7 to point 4 can occur. This flow is unwanted and can be referred as a reverse heat flow that can cover the bearing temperature rise signature caused by bearing heat generation system.

The chosen heaters are also re-checked using finite element modeling using ANSYS software. The results shown in Figure 7.4 assure that the reverse heat flow is prevented and the direction of the heat flow is always from the bearing outer race to the lower housing. This is depicted in Figure 6.3 and can be concluded by observing the color code which indicates that the heat flow in z direction, the vertical direction, is negative at the neighborhood of bearing race and positive at the neighborhood of the surface heater.

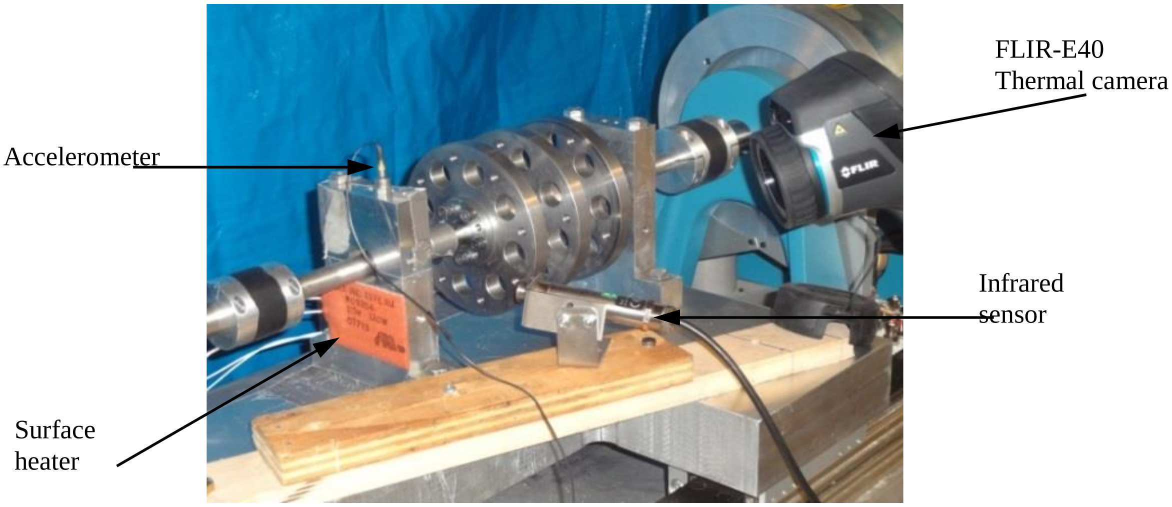


Figure 7.2 Bearing test rig with a surface heater installed

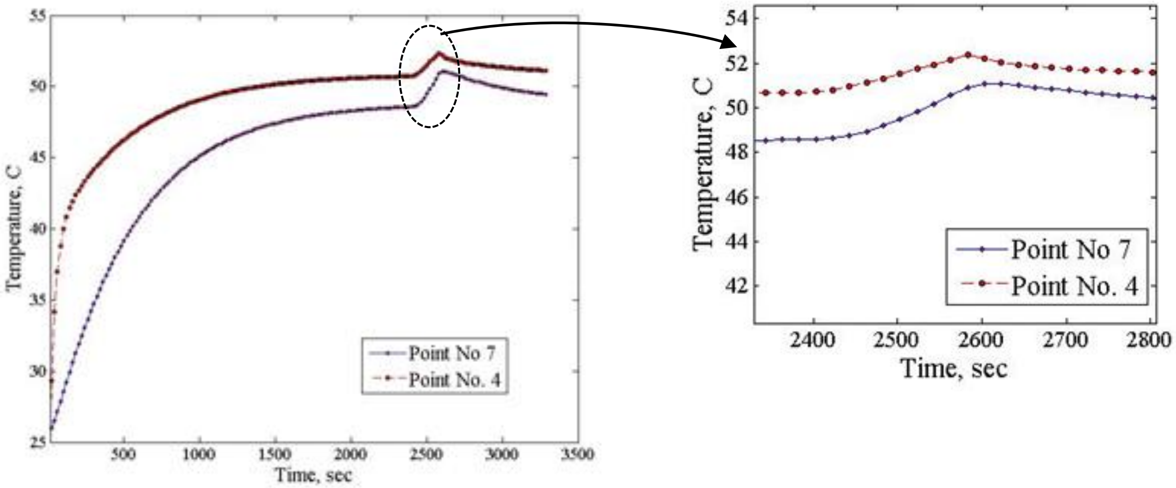


Figure 7.3 Temperature rise curves calculated at points No. 7 and point No. 4 from Figure 6.1, using the lumped mass method model

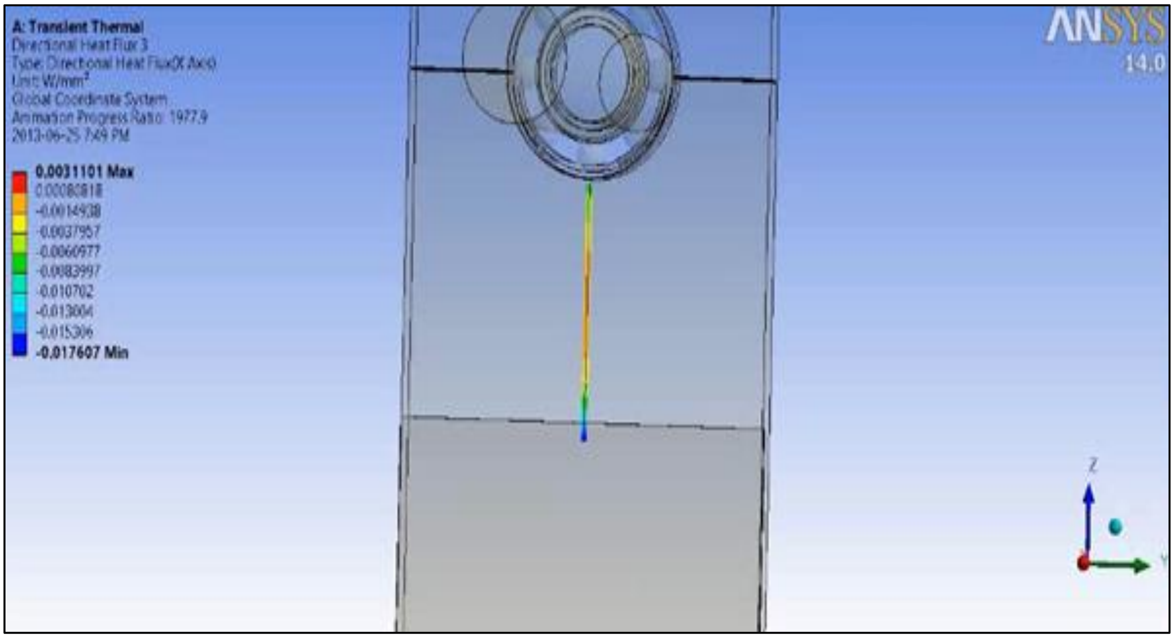


Figure 7.4 Finite element analysis results using ANSYS at the end of the energizing time that shows the heat flow direction is in negative z direction

7.2 Experimental Results

The experimental setup shown in Figure 7.2 is used in testing the capability of the proposed technique in bearing fault detection. Firstly the temperature curve for a healthy bearing is recorded. The healthy bearing is let to run until it reaches thermal equilibrium condition, then the surface heater is energized for a 3 minutes. The temperature curves are recorded for the healthy bearing operation at initial transient, equilibrium and stimulated transient condition. These records are kept and will be used as reference bearing temperature curves. Then the faulty bearing with outer race fault will let to run at the same operating conditions, same load and same speed. The temperature rise curves are recorded for the three stages as performed with the healthy bearing. The recorded temperature rise curves are compared as shown in Figure 7.5. The comparison shows that the temperature rise curve for healthy and faulty bearing will be different when the bearing in either the initial transient condition or the stimulated transient condition. The results also show that the steepness of the curves for the stimulated transient condition is lower than that for the initial transient condition. This is due to the change of thermal properties at different temperature range. However, the detection of the bearing fault depending on difference in temperature rise curves is possible and can be a candidate for more development especially for bearing testing under controlled conditions and in laboratories. The stimulation method is necessary only after reaching thermal equilibrium state. Afterwards, it can be done periodically. The time intervals for each system or operationg condition is situation-dependent.

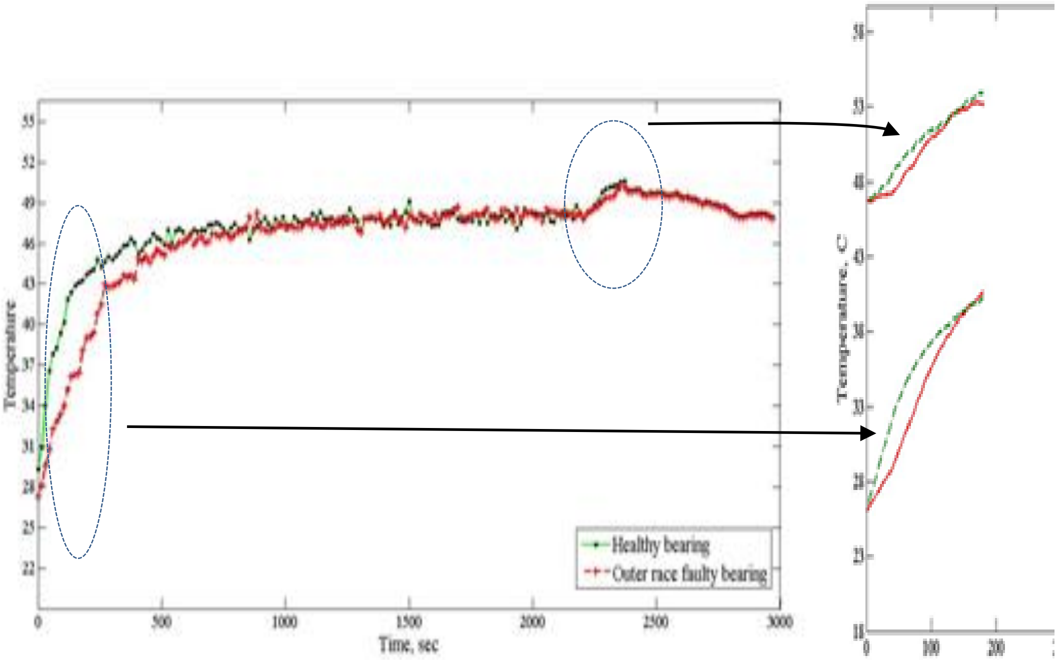


Figure 7.5 Experimental results for both the initial transient and stimulated transient conditions for both healthy and faulty bearings

Chapter 8

Summary, Conclusion and Future Work

8.1 Summary

In the current study the exploitation of the progress in two rarely-coupled fields for rolling element bearing health assessment is attempted. The first field is tribology with one of its branches in rolling contact friction and the generated heat. The second field is the bearing condition monitoring and the most common vibration analysis methods. The coupling of the two fields is manifested by modeling the bearing friction and calculating the resulted bearing temperature for different bearing conditions then using the results to compare with the results of vibration-based bearing health assessment methods.

In both Chapter 3 and Chapter 4, the obstacles that hinder the effectiveness of temperature monitoring methods for rolling element bearing early fault detection are revealed. In Chapter 3, the details about the possible bearing heat sources and bearing temperature calculation methods including the two of the most popular bearing heat calculation methods are presented. In Chapter 4 a three-module analytical model for bearing temperature calculation is proposed. The result of the first two modules, i.e., bearing total heat generation rate, is compared to the bearing heat that is calculated by the two existing methods, Palmgren and SKF that are previously discussed in Chapter 3. Thence and by adding the third module to the previous two modules, the bearing transient temperature is calculated and compared to the measured temperature using a bearing temperature-vibration test rig that is specifically built for this purpose.

In Chapter 5, the amplitude demodulation or envelope analysis was chosen to accomplish the vibration analysis part. The method is enhanced through two steps: the optimum selection of the band pass filter using the SK-STFT method and the in-band signal enhancement using both LPF and the new ATSA method. It is shown that by incorporating the two steps in the basic envelope analysis technique, clearer fault detection is attained.

In Chapter 6 an attempt to evaluate the bearing health assessment using the two different trends, i.e., temperature-based and vibration-based methods, inspired a new classification method of bearing faults. The experimental tests are designed to show the advantages and the pitfalls of each trend in accordance with the fault type and the test conditions. The results of bearing temperature-based fault detection method showed that the transient temperature curve can be considered as a health assessment indicator that is useful in both of the two types of faults while the bearing equilibrium temperature is not so indicative of bearing condition. The application of vibration-based method showed a problem of bearing health assessment when the fault is matured and when the fault is of the NIF type.

In Chapter 7, to enhance the temperature-based method in the case of fault occurrence after reaching the thermal equilibrium condition, a new technique for stimulating the transient condition is proposed. The results showed that by stimulating the transient condition of bearing thermal system, the deviation between the healthy and faulty bearing transient temperature becomes noticeable and can be used for health assessment purpose.

8.2 Conclusions

The main accomplishments and findings of this thesis work are as follows:

1. An analytic solution for bearing heat generation rate is accomplished by dividing the heat sources into three parts: the dry friction which is analyzed using the Hertzian solution for contact area stress distributions at different loading conditions, the hydrodynamic friction using the simplified Reynolds equation and the Hamrock and Dowson empirical formula (Hamrock, et al., 1977) for the lubricant film thickness and the lubricant viscous drag using the formula proposed by Harris (Harris, et al., 2006). The heat generation results obtained using the proposed model showed that the obtained heat generation rates is in between the results of two empirical models, Palmgren (Palmgren, 1959) and SKF (SKF, 2003) models in addition to a more sensitivity to load variations as depicted in Figures 4.4 and 4.5.
2. The proposed bearing temperature model showed a closer agreement with that of the experimental tests compared to the results of the two empirical methods either when considering constant lubricant properties or when using the adaptive scheme for bearing transient temperature calculation. The results of the non-adaptive and the adaptive bearing temperature algorithm, depicted in Figures 4.10 and 4.11, showed that: for the non-adaptive algorithm the results of the three methods deviate from the experimental results. On the other hand, using the adaptive algorithm that counts for lubricant property changes with the temperature rise reduces the difference between the three models and the experimental results. The enhancement of the proposed model by applying the adaptive algorithm is greater than the enhancement of the results by the two other methods.
3. An enhanced envelope analysis is proposed based on two steps: the optimum selection of the envelope window using SK-STFT-based method, and then the enhancement of the in-band signal by using LPF for deterministic noise removal and a new proposed method, Autonomous Time Synchronous Averaging ATSA. The ATSA method re-aligns the signal by aligning the points of maximum Teager operator in non-overlapping segments of the signal. The two steps are used to study the effect of both rolling element slippage and high interference caused by a gear mesh vibration on bearing early fault detection. The simulation has shown that the bearing slippage can reduce the SNR for the resulted envelope spectrum can be reduced from 15 dB to 1.7 dB for $\pm 2\%$ slippage ratio and to -20 dB for $\pm 5\%$ slippage ratio. Although using the TSA was successful to raise the SNR

for the first case to 6.3 dB, for higher slippage ratios it did not succeed to reveal the fault as the SNR was raised from -20 dB to -7.2 dB which is not enough for fault detection. On the other hand, using the proposed ATSA algorithm could increase the SNR for the first case to 14 dB and for the higher slippage ratio to 7.8 dB that is sufficient in most cases for detecting bearing faults in the envelope spectrum. The combined effect of rolling element slippage and the presence of a strong interfering signal (i.e., a gear meshing vibration signal) was also studied using simulated signals. The simulated signal with the combined effect that has a SIR of -21 dB at $\pm 2\%$ slippage ratio and of -53 dB at $\pm 5\%$ was enhanced to 11 dB and 1.45 dB using LPF and TSA before envelope analysis. In comparison, the application of the LPF and the ATSA before the envelope analysis raised the SIR for the same simulated signals to 32 dB and 20 dB, respectively.

4. Classifying bearing faults into IFT, faults with strong impulsive vibration signal and NIF, faults with weak or unnoticeable impulses, is found to be useful in explaining the importance of using transient temperature monitoring method in addition to the vibration-based monitoring methods for better early bearing fault detection. The application of both methods on real faults of the two types showed that the transient temperature monitoring method was successful in detecting both types of bearing faults while the vibration-based method was not successful for the NIF and the IFT mature faults. The tests also showed that the temperature-based method was successful in fault detection only by comparing the transient temperature curves while the equilibrium temperatures were shown to be inadequate for the detection of any bearing faults of the two types.
5. As it is concluded from the results of all temperature-based monitoring methods that the detection of bearing faulty condition is only possible during the thermal transient condition, stimulating the transient condition is accomplished using a transient stimulation system, TSS system. The TSS system was designed using the lumped mass system and finite element modeling to find the power of the surface heaters and the energizing time for stimulating the transient condition for the bearing without the domination of the heat flow from the surface heaters on the heat generation signature of the bearing. The results of the TSS system indicated that the faulty bearing condition has

a detectable signature compared to that for the healthy bearing during the stimulated transient condition as well as that was found at the initial transient condition.

6. The steps and results listed above show that the combination between temperature and vibration methods can lead to more efficient bearing health assessment systems. Unlike the vibration-based methods that are affected by the development of the fault shape and the resulted weaken signal, the temperature-based methods are not affected by the impulsiveness of the signal and can still detect the fault after it becomes blunt. However, the temperature-based methods are sensitive to any change of the tribology system that does require accurate measurement devices for good indexing of tribology system differentiation. For the current available temperature devices it can be said that this technique is better suited for controlled environment and laboratory testing than in-field testing.

8.3 Future Work

Temperature-based monitoring methods for mechanical systems were hindered for long time due to the rapid variations of some mechanical system parameters and slow response of ordinary temperature monitoring devices, as well as the low accuracy of these devices that make the accuracy error in many cases greater than the change caused by the mechanical system anomalies. These two obstacles were partially overcome in the current study by using a high response temperature monitoring device, a thermal camera and an infrared sensor, and by monitoring the system at the transient condition, either real or stimulated. However, the success of the methods in the current study depends on how well the test conditions are controlled in a laboratory environment. The future work using the temperature-based method should focus on the following points:

1. The development of more advanced temperature monitoring devices that have higher responsiveness and lower error margins.
2. The study of the bearing transient temperature under severe conditions such as a fast change of environment temperature or under windy conditions.
3. More explanation for the temperature rise curve for each change in rolling contact including the effects of fault shape, fault dimensions and fault location.

For the vibration-based methods, the following enhancement could be attempted:

1. The window selection algorithm can be examined with different time-frequency domain transformations for faster calculation for a wider range of window sizes and window overlapping ratios.
2. The in-band bearing signal enhancement can be tested using different discriminator other than the Teager operator searching for better enhancement for bearing signal envelope spectra.
3. The change of vibration signal impulsiveness that was measured as a change in the power of the envelope spectrum at fault frequencies can be used as a measure for the fault maturity and the URL according to the combination of the stored history of the bearing vibration and the results of the TTS system application.

Another interesting direction for the future work could be sensor fusion, i. e., integration of the information from both the temperature and vibration sensors to obtain a fused index for on-line decision making.

References

Antoni, J. (2006). The Spectral Kurtosis:a Useful Tool for Characterising Nonstationary Signals. *Mechanical Systems and Signal Processing, 20*(2), 308-331.

Antoni, J., & Randall, R. B. (2001). Optimisation of SANC for Separating Gear and Bearing Signals. In A. G. Starr, & R. B. Rao (Ed.), *Proceeding of the 14th International Congress Condition Monitoring and Diagnosic Management,* (pp. 89-97). Manchester, UK.

Antoni, J., & Randall, R. B. (2004). The Spectral Kurtosis: Application to the Surveillance and Diagnostics of Rotating Machines. *Mechanical Systems and Signal Processing, 20*(2), 308-331.

Antoni, J., & Randall, R. B. (2004). Unsupervised Noise Cancellation for Vibration Signals: part II—a Novel Frequency-Domain Algorithm. *Mechanical systems and signal processing, 18*(3), 103-117.

Antoni, J., Bonnardot, F., Raad, A., & El Badaoui, M. (2004). Cyclostationary Modelling of Rotating Machine Vibration Signal. *Mechanical Systems and Signal Processing, 18*(6), 1285-1314.

Azad, D., & Ramji, K. (2012, August). Identification of Bearing Assembly Defects Using Finite Element Analysis and Condition Monitoring Techniques. *International Journal of Engineering Research & Technology (IJERT), 1*(6), 1-13.

Balderston, H. L. (1969, June). The Detection of Incipient Failures in Bearings. *Material Evaluation Journal, 27*(4), 121-128.

Bechhoefer, E., & Praneet, M. (September 27-October 1, 2009). Bearing Envelope Analysis

Window Selection. In B. SAHA, & K. Goebel (Ed.), *Annual Conference of the Prognostics and Health Management Society*, (pp. 1-7). San Diego, CA.

Bechhoefer, E., Kingsley, M., & Menon, P. (2011). Bearing Envelope Analysis Window Selection using Spectral Kurtosis Techniques. In S. Shieh (Ed.), *Proceeding of 2011 IEEE Conference on Prognostics and Health Management (PHM)* (pp. 1-6). Denever, CO: IEEE.

Becker, A. A. (1992). *The Boundary Element Method in Engineering, a Complete Course.* New York, USA: McGraw Hill International.

Boffy, H., Baietto, M. C., Sainsot, P., & Lubrecht, A. A. (2012, Feb.). Detailed Modelling of a Moving Heat Source Using Multigrid Methods. *Tribology International, 46*(1), 279-287.

Bozchalooi, I. S., & Liang, M. (2007). A Smoothness Index Guided Approach to Wavelet Parameter Selection in Signal De-Noising and Fault Detection. *Journal of Sound and Vibration, 308*(1-2), 246-267.

Bozchalooi, I. S., & Liang, M. (2008). A Joint Resonance Frequency Estimation and In-Band Noise Reduction Method for Enhancing the Detectability of Bearing Fault Signals. *Mechanical Systems and Signal Processing, 22*(4), 915-933.

Bozchalooi, I., & Liang, M. (2010). Teager Energy Operator for Multi-Modulation Extraction and its Application for Gearbox Fault Detection. *Smart Materials and Structures, 19*(7), 1-18.

Branch, N. A., Arakere, N. K., Svendsen, V., & Forester, N. H. (2009). Stress Field Evolution in a Ball Bearing Raceway Fatigue Spall. *Journal of ASTM International*, *7*(2), 1-40.

Braun, S. (2011). The Synchronous (Time Domain) Average Revisited. *Mechanical Systems and*

Signal Processing, 25(4), 1087-1102.

Braun, S., & Datner, B. (1979). Analysis of Roller/Ball Bearing Vibrations. *ASME Journal of Mechanical Design, 101*, 118-125.

Brewe, D. E., & Hamrock, B. J. (1977, October). Simplified Solution for Elliptical Contact Deformation Between two Elastic Solids. *Journal of Lubrication Technology, 99*(4), 485-487.

Bush, A., Gibson, R., & Thomas, T. (1975). The Elastic Contact of a Rough Surface. *Wear, 35*(1), 87-111.

Cameron, A. (1985). Righting a 40-Year-Old Wrong; A.M. Ertel–The True Author of Grubin's EHL Solution. *Tribology International, 18*(2), 65-136.

Capdevielle, V., Serveiere, C., & Lacoume, J.-L. (September 10-13, 1996). Blind Separation of Wide-Band Sources: Application to Rotating Machines Signals. In G. Ramponi, L. Sicuranza, S. Carrato, & S. Marsi (Ed.), *Proceedings of the Eighth European Signal Processing Conference, Vol. 3*, (pp. 2085-2088). Trieste, Italy.

Croes, J., & Iqbal, S. (2009). *www.estomad.org.* Retrieved from www.ESTOMAD.org: www.estomad.org/documenten

Damiens, B., Venner, C. H., Cann, P. M., & Lubrecht, A. A. (2004, January). Starved Lubrication of Elliptical EHD Contacts. *Journal of Tribology, 126*(1), 105-111.

Dwyer, R. (1983). Detection of Non-Gaussian Signals by Frequency Domain Kurtosis Estimation. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing 8* (pp. 607-610). Boston, Massachussets: IEEE.

Flir Incorporation. (2011). Data Sheet for Flir E40 Thermal Camera.

Fridman, H. D., & Levesque, P. (1959, june). Reduction of Static Friction by Sonic Vibrations. *Journal of Applied Physics, 30*(10), 1572-1575.

Gardner, W. A., & Franks, L. E. (1975, January). Characterization of Cyclostationary Random Signal Processes. *IEEE transaction on Information Theory, 21*(1), 4-14.

Godfrey, D. (1967). Vibration Reduces Metal to Metal Contact and Causes an Apparent Reduction in Friction. *ASLE Transaction, 10*(2), 183-192.

Greenwood, J. A., & Tabor, D. (1958). The Friction of Hard Sliders on Lubricated Rubber:The Importance of Deformation Losses. *Proceedings of the Physical Society, 71*(6), 989-1001.

Greenwood, J., & Williamson, J. (1966). Contact of Nominally Flat Surfaces. *Proceeding of Royal Society London, 295*(1442), 300-319.

Hamrock, B. J. (1975). Ball Motion and Sliding Friction in Arched Outer Race Ball Bearing. *Journal of Tribology, 97*(2), 202-210.

Hamrock, B. J., & Dowson, D. (1976, July). Isothermal Elastohydrodyanamic lubrication of point contacts: Part-II Ellipticity parameter results. *Journal of Lubrication Technology, 98*(3), 375-381.

Hamrock, B. J., & Dowson, D. (1977, April). Isothermal Elastohydrodynamic Lubrication of Point Contacts. *Journal of Lubrication Technology, 99*(1), 264-276.

Hamrock, B. J., Schmid, S. R., & Jacobson, B. O. (2004). Introduction. In *Fundamentals of Fluid Film Lubrication* (2 ed., p. 7). New York, USA: Marcel Dekker Inc.

Harris , T., & Mindel, M. H. (1973). Rolling Element Bearing Dynamics. *Wear, 23*(3), 311-337.

Harris, T. (1971, January). An Analytical Method to Predict Skidding in Thrust Loaded, Angular-Contact Ball Bearings. *ASME Transactions, J. Lubrication Technology, 93*(1), 17-24.

Harris, T. A., & Kotzalas, M. N. (2007). *Advanced Concepts of Bearing Technology* (5 ed.). Boca Raton, FL, USA: Taylor and Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Contact Stress and Deformation. In *Essential Concepts of Bearing Technology* (5 ed., pp. 101-133). Boca Raton, FL, USA: Taylor and Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Friction in Rolling Element-Raceway Contacts. In *Advanced Concepts of Bearing Technology* (5 ed., pp. 143-168). Boca Raton, FL, USA: Taylor & Francis.

Harris, T. A., & Kotzalas, M. N. (2007). High Speed Operation: Ball and Roller Dynamic Loads and Bearing Internal Load Distribution. In *Advanced Concepts of Bearing Technology* (5 ed., pp. 79-122). Boca Raton, FL, USA: Taylor & Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Lubricant Films in Rolling Element-Raceway Contact. In *Advanced Concepts of Bearing Technology* (5 ed., pp. 113-117). Boca Raton, FL: Taylor& Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Lubricant Films in Rolling Element-Raceway Contact. In *Advanced Concepts of Bearing Technology* (5 ed., pp. 123-142). Boca Raton, FL, USA: Taylor and Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Rolling bearing temperatures. In *Advanced Concepts of Bearing Technology* (pp. 207-225). Boca Raton, FL: Taylor & Francis.

Harris, T. A., & Kotzalas, M. N. (2007). Rolling Bearings Types and Applications. In *Essential*

Concepts of Bearing Technology (5 ed., pp. 1-4). Boca Raton, FL, USA: Taylor and Francis.

Henao-Sepulveda, J. A., Toledo-Quinones, M., & Jia, Y. (2005). Contactless Monitoring of Ball Bearing Temperature. In E. M. Petriu (Ed.), *Proceedings of the IEEE Instrumentation and Measurement Technology Conference* (pp. 1571-1573). Ottawa, Canada.

Hess, D. P., & Soom, A. (1991, January). Normal Vibrations and Friction under Harmonic Loads: Part I- Hertzian Contacts. *Journal of Tribology, 113*(1), 80-86.

Hess, D. P., & Soom, A. (1991, Januray). Normal Vibrations and Friction under Harmonic Loads: Part II-Rough Planar Contacts. *Journal of Tribology, 113*(1), 87-92.

Ho, D., & Randall, R. B. (2000). Optimisation of Bearing Diagnosic Techniques using Simulated and Actual Fault Signals. *Mechanical Systems and Signal Processing, 14*(5), 763-788.

Inceropera, DeWitt, Bergman, & Lavine. (2007). *Fundamentals of Heat and Mass Transfer* (6 ed.). New York, USA: John Wiley and Sons.

Inceropera, DeWitt, Bergman, & Lavine. (2007). Transient Conduction. In *Fundamentals of Heat and Mass Transfer* (6 ed., pp. 260-261). New York, USA: John Wiley and Sons.

Jeng, Y.-R., & Haung, P.-Y. (2003). Prediction for Temperature Rise for Ball Bearing. *Tribology Transaction, 46*(1), 49-56.

Jeng, Y.-R., & Wang, P.-Y. (2003, April). An Elliptical Microcontact Model Considering Elastic, Elastoplastic and Plastic Deformation. (L. Chang, Ed.) *ASME Journal of Tribology, 125*(2), 232-240.

Johnson, K. L. (1985). *Contact Mechanics.* Cambridge, UK: Cambridge University Press.

Johnson, M. (2009, April). Lubricant Application:Grease Volumes and Frequencies. *Tribology and Lubrication Technology, 2,* 1-7.

Jones, A. (1959). Ball Motion and Sliding Friction in Ball Bearing. *ASME Transactions Journal of Basic Engineering, 81*(7), 1-12.

Kim, W., Jinju, S., & Hong, D. (2012). Infrared Thermographic Inspection of Ball Bearing; Condition Monitoring. In R. Diederichs (Ed.), *Proceedings of the 18th World Conference of Non Destructive Testing*, (pp. 1-4). Durban, South Africa.

Kurfess, T. R., Billington, S., & Liang, S. Y. (2006). Advanced Diagnostic and Prognostic Techniques for Rolling Elements Bearing. In Lihui Wang, & Robert X. Gao, *Condition Monitoring and Control for Intelligent Manufacturing* (pp. 153-181). London, UK: Springer.

Lienhard IV, J. H., & Lienhard V, J. H. (2003). *A Heat Transfer Text Book* (3 ed.). Cambridge, Massachustess, USA: Phlogiston Press.

Lubrecht, A. A., ten Napel, W. E., & Bosma, R. (1987). Multigrid, an alternative method of solution for Two-Dimensional Elastohydrodynamically Lubricated Point Contact Calculations. *Journal of Tribology, 109*(3), 437-443.

Lugt, P. M., & Morales-Espejel, G. E. (2011). A Review of Elasto-Hydrodynamic Lubrication Theory. *Tribology Transaction, 54*(3), 470-496.

Lumax. (2011, Feb 17). Lumax Lubrication Equipment: Material Safety Data Sheet, MC2048BT.

Maldague, X. (2001). *Theory and Practice of Infrared Technology for Non Destructive Testing,.* Laval, QC: John-Wiley &. Sons,.

Marble, S., & Tow, D. (4-11 March, 2006). Bearing Health Monitoring and Life Extension in Satellite Momentum/Reaction Wheels. *Proceedings of the 11th IEEE Aerospace Conference*, (pp. 1-7). Big Sky, Montana.

McFadden, P. D., & Smith, J. D. (1984). Model for the Vibration Produced by a Single Point Defect in Rolling Element Bearing. *Journal of Sound and Vibration, 96*(1), 69-82.

McFadden, P. D., & Smith, J. D. (1984). Vibration Monitoring of Rolling Element Bearings by the High Frequency Resonance Technique- a Review. *Tribology International, 17*(1), 3-10.

McFadden, P. D., & Toozhy, M. M. (2000). Application of Synchronous averaging to Vibration Monitoring of Rolling Element Bearings. *Mechanical Systems and Signal Processing, 14*(6), 891-906.

Mizuta, K., Inoue, T., Haung, S., Ueda, K., & Omokawa, H. (2003). Heat Transfer Characteristics Between Inner and Outer Rings of an Angular Ball Bearing. *Heat Transfer—Asian Research, 32*(1), 42-57.

Mourier, L., Mazuyer, D., Lubrecht, A. A., & Donnet, C. (2006). Transient Increase of Film Thickness in Micro-Textured EHL Contact. *Tribology International, 39*(12), 1745-1756.

Nayfeh, A. H., & Mook, T. D. (1979). *Nonlinear Oscillation* (1 ed.). Virginia, USA: John Wiley & Sons.

Nijenbanning, G., Venner, C. H., & Moes, H. (1994). Film thickness in Elastohydrodynamically Lubricated Elliptic Contact. *Wear, 176*(2), 217-229.

Omega Incorporation. (2013). OS210 series Infrared Sensor User Guide.

Otonnello, C., & Pagnan, S. (1994). Modified Frequency Domain Kurtosis for Signal Processing.

Electronics Letters, 30(14), 1117-1118.

Palacios, J. M., & Palacios, M. P. (1984, June). Rheological Properties of greases in EHD Contacts. *Tribology International, 17*(3), 167-171.

Palmgren, A. (1959). *Ball and Roller Bearing Engineering* (3rd ed.). Burbank, Philladelphia: SKF Industries.

Patrick, R., Smith, M. J., Zhang, B., Byington, C. S., Vachtsevanos, G. J., & Rosario, R. D. (March 7-14, 2009, January 6). Diagnostic Enhancements for Air Vehicle HUMS to Increase Prognostic System Effectiveness. In T. Vladimirova (Ed.), *Proceedings of the 14th IEEE Aerospace conference* (pp. 1-12). Big Sky, Montana: IEEE.

Pitts, D. R., & Sissom, E. L. (1997). *Theory and Problems of Heat Transfer* (2 ed.). New York, USA: McGraw Hill.

Plotinkov, Y. A., & Winfree, W. P. (25-30 July, 1999). Temporal Treatment of a Thermal Response for Defect Depth Estimation. In C. P. Burgess, & R. C. Myers (Ed.), *Proceedings of American Institute of Physics conference #509*, (pp. 587-595). Montreal, Canada.

Popov, E. P. (1990). Axial Strains and Deformation in Bars. In *Engineering Mechanics of Solids* (pp. 96-99). New Jersey, USA: Prentice-Hall Inc.

Popov, V. L. (2010). *Contact Mechanics and Friction; Physical Principles and Applications.* Berlin, Germany: Springer.

Proakis, J. G., & Manolakis, D. G. (1996). Power Spectrum Estimation. In *Digital Signal Processing, Principles, Algorithms and Applications* (pp. 925-930). New Jersy, USA: Prentice Hall.

Qiu, H., Lee, J., in, J., & Yu, G. (2003). Robust Performance Degradation Assessment Methods for Enhanced Rolling Element Bearing Prognostics. *Advanced Engineering Informatics, 17*(3-4), 127-140.

Qiu, H., Lee, J., in, J., & Yu, G. (2006). Wavelet Filter-Based Weak Signature Detection Method and its Application on Rolling Element Bearing Prognostics. *Journal of Sound and Vibration, 289*(4-5), 1066-1090.

Randall, R. B. (2011). *Vibration-Based Condition Monitoring.* West Sussex, UK: A John Wiley and sons.

Randall, R. B., & Antoni, J. (2011). Rolling Element Bearing Diagnostics—A Tutorial. *Mechanical Systems and Signal Processing, 25*(2), 485-520.

Sakamoto, S., Sakamoto, T., Taguchi, E., & Nakao, K. (2012, October). Effects of Vibration on Frictional Resistance during Angioplasty Device Insertion: Possible Application of a Newly Designed Support System to Complex Coronary Intervention. *Journal of the American College of Cardiology, 60*(17), B128-B129.

Sakamoto, T. (1987). Normal Displacement and Dynamic Friction Characteristics in a Stick-Slip Process. *Tribology International, 20*(1), 25-31.

Sakamoto, T., & Tsukizoe, T. (1978). Metal Transfer in the Frictional Contact of a Rough Hard Surface. *Wear, 47*(2), 301-313.

Sawalhi, N., & Randall, R. B. (2004). The Application of Spectral Kurtosis to Bearing Diagnostics. In D. J. Mee, R. J. Hooker, & I. Hillock (Ed.), *Proceedings of ACOUSTICS conference,* (pp. 393-398). Gold Coast, Australia.

Sawalhi, N., Randall, R. B., & Endo, H. (2007). The Enhancement of Fault Detection and

Diagnosis in Rolling Element Bearings Using Minimum Entropy Deconvolution Combined with Spectral Kurtosis. *Mechanical system and signal processing, 21*(6), 2616-2633.

Schipper, D. J., Vroegop, P. H., de Gee, A. J., & Bosma, R. (1990, April). Micro-EHL in Lubricated Concenterated Contacts. *Journal of Tribology, 112*(2), 392-397.

Schneider, Y., Zahn, S., Schindler, C., & Rohm, H. (2009). Ultrasonic Excitation Affects Friction Interactions between Food Materials and Cutting Tools. *Ultrasonics, 49*(6-7), 588–593.

Scott, S., Kovacs, A., Gupta, L., Katz, J., Sadeghi, F., & Peroulis, D. (2011). Wireless Temperature Microsensors Integrated on Bearings for Health Monitoring Applications. In K. F. Bohringer, & L. Liwei (Ed.), *2011 IEEE 24th International Conference on Micro Electro Mechanical Systems (MEMS)* (pp. 660-663). Cancun, Mexico.: IEEE.

Seireg, A. A. (1998). *Friction and Lubrication in Mechanical Design.* New York: Marcell and Dekker.

Seo, J., Yoon, H., Ha, H., Hong, D., & Kim, W. (2011). Inferarred Thermographic Diagnosis Mechanism for Fault Detection of Ball Bearing under Dynamic Loading Conditions. *Advanced materials research, 295-297*, 1544-1547.

Sheen, Y.-T., & Liu, Y.-H. (2012). A Quantified Index for Bearing Vibration Analysis Based on the Resonance Modes of Mechanical System. *Journal of Intelligent Manufacturing, 23*(2), 189-203.

SKF. (2003, June). SKF General Catalog, 5000E. SKF Incorporation. Retrieved from www.skf.com.

Stachowiak, G. W., & Batchelor, A. W. (2014). *Engineering Tribology* (4 ed.). Oxford, UK: Butterworth Heinemann.

Stipanovic, A. J. (2003). Hydrocarbon Base Oil Chemistry. In G. E. Totten, S. R. Westbrook, & R. J. Shah (Eds.), *Fuels and Lubricants Handbook; Technology, Properties, Performance and Testing* (pp. 169-184). West Conshohocken, PA, USA: ASTM International.

Stolarski, T. A. (1990). *Tribology in Machine Design.* Oxford, UK: Butterworth-Heinemann.

Stolarski, T. A., & Tobe, S. (2000). Rolling Contacts. London, UK: Professional Engineering.

Streeter, V. L. (1962). *Viscous Effect-Fluid Resistance* (3 ed.). New York, USA: McGraw-Hill.

Takabi, J., & Khonsari, M. M. (2013). Experimental Testing and Thermal Analysis of Ball Bearings. *Tribology International, 60*, 93-103.

Tea, P. L., Castelli, V., Murdock, J. W., & Meirovetch, L. (2007). Mechanics of Solids and Fluids. In E. A. Avallone, T. I. Baumeister, & A. M. Sadegh (Eds.), *Mark's Standard Handbook for Mechanical Engineers* (pp. 20-29). New York, USA: McGraw Hill.

Teager, H., & Teager, S. (1990). Evidence for Nonlinear Sound Production Mechanism in the Vocal Tract Speech Production and Speech Modeling. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech Production and Speech Modeling* (Vol. 55, pp. 241-261). Netherland: Springer.

Tolstoi, D. M. (1967). Significance of the Degree of Freedom and Natural Normal Vibrations in Contact Friction. *Wear, 10*(3), 199-213.

Tolstoi, D. M., Borisova, G. A., & Grigorova, S. R. (1973). Friction Reduction by Perpendicular Oscillation. *Soviet Physics-Doklady, 17*(9), 907-909.

Trachman, E., & Cheng, H. (1972). Thermal and non-Newtonian Effects on Traction in Elastohydrodynaimc Contacts. *Proceedings of the Institute of Mechanical Engineers 2nd symposium on Elastohydrodynamic lubrication*, (pp. 142-148). Leeds.

US Enviromental Protection Agency. (2011). *Screening-Level Hazard Characterization Waxes and Related Materials Category*. Retrieved from www.epa.gov.

Venner, C. H., & Lubrecht, A. A. (2000). Multigrid. In D. Dowson (Ed.), *MultiLevel Methods in Lubrication* (1 ed., pp. 57-97). Amesterdam, Netherlands: Elsevier.

Wang, Y., & Liang, M. (2011). An Adaptive SK Technique and its Application for Fault Detection of Rolling Element Bearings. *Mechanical Systems and Signal Processing, 25*(5), 1750-1764.

Watson, M. J., Carl S. Byngton, & Behbahani, A. (2007). Very High Frequency Monitoring System for Engine Gearbox. *SAE Technical Paper 2007-01-3878*.

Winfield, L., Glassmire, J., Colgate, E. J., & Peshkin, M. (2007). T-PaD: Tactile Pattern Display through Variable Friction Reduction. In H. Iwata (Ed.), *proceedins of the World Haptics Conference*, (pp. 421-426). Tsukuba, Japan.

Cao, Y. (2006). Modeling of High-Speed Machine-Tool Spindle Systems. *PhD Thesis*. BC, Canada: Mechanical Engineering Department, University of British Colombia.

Zhu, D., & Wang, Q. J. (2011, October). ElastoHydroDynamic Lubrication:a Gateway to Interfacial Mechanics, Review and Prospect. *Journal of Tribology, 133*(4), 1-14.

APPENDIX

Lubricant Rheology

The used lubricant in bearing under test is LUMAX-1911 which is a lithium based grease with a base oil which is a mixture of two or more highly refined petroleum oils such as; CAS 64741-88-4 and CAS 64741-89-5 or similar pairs of oils as indicated in (Lumax, 2011). The mixture has a medium carbon number, from C20 to C40, which is in the middle between light and heavy refined oils and has an index of viscosity, VI, equals 150. The approximate measurement for the viscosity of this type of base oil at ambient temperature, 20 °C is 92 cSt, at 40 °C is 16.6 cSt and 4 cSt at 100 °C (Stipanovic, 2003). Because this mixture contains oils with different viscosities and without a precise percentage of content, a correction of the effective viscosity for the mixture will be performed based on the measured friction heat. that builds up the base oil for the lubricating grease has a minimum kinematic viscosity of 19 cSt at 40° Celsius (US Enviromental Protection Agency, 2011). The properties of the grease, used in the bearing under test, such as viscosity and thermal conductivity are mainly returned to that of the base oil. These properties are affected by three different factors; temperature, pressure and shear rate. Most of the studies done so far to find the effective viscosity of the lubricant at certain operating conditions are based on empirical studies. Combining the effect of the three effectors is a very complex task. The empirical studies showed that the viscosity of the lubricant decreases with the increase of the operating temperature and increases with the increase of the applied pressure. These two results were formulized by several researchers, e. g. Walther and Vogel for viscosity-temperature formula, Barus and Roelands for viscosity-pressure relationship.

The third factor that affects the lubricant effective viscosity is the shear rate which has different effect on different types of lubricants. When the shear rate gets higher than some limit for a specific fluid, it can cause either an increase or a decrease of the fluid effective viscosity. The grease is one of the second type for which when the shear rate exceeds some limit the effective viscosity starts to decrease till reaching a steady viscosity value.

The Vogels equation will be used to find and graph temperature–viscosity relationship for the base oil.

A.1 Temperature Effect on Lubricant Properties

The lubricant viscosity change exponentially with the increase of the temperature this is described in Vogels Equation (A.1) while both thermal conductivity and specific heat are linearly proportional to the operating temperature. The specific heat of a lubricant increases with the increase in operating temperature while thermal conductivity decreases with temperature increase. Equations (A.2) and (A.3) describe approximate formulas to estimate specific heat and thermal conductivity for the lubricant at specific temperature (Stachowiak, et al., 2014).

$$\eta = ae^{b/(T-c)} \tag{A-1}$$

$$\sigma = \left(1.63 + 0.0034T\right) / s^{0.5} \tag{A-2}$$

$$k = (0.12/s) \times \left(1 - 1.667 \times 10^{-4} T\right) \tag{A-3}$$

where η is the effective viscosity, a, b and c are three constants and T is the operating temperature in degrees Celsius. σ is the specific heat (KJ/Kg. K), k is the thermal conductivity in W/mK and s is the specific gravity at 15.6 °C which equals 0.88. By knowing the lubricant viscosity at three different temperatures, the above formula is solved to find the three constants. The resulting graph of viscosity-temperature relationship is shown in Figure A.1.

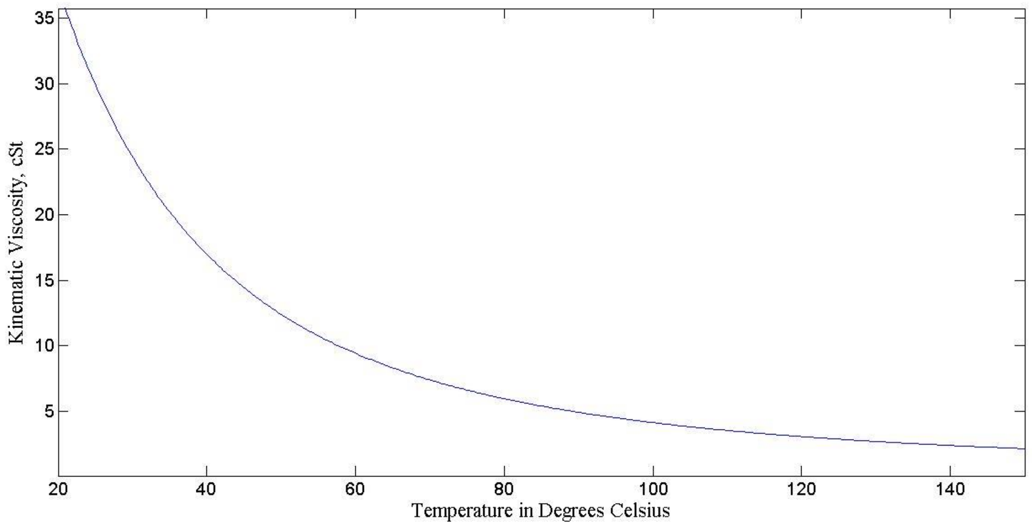


Figure A.1 Lubricant viscosity versus operating temperature using Vogels equation.

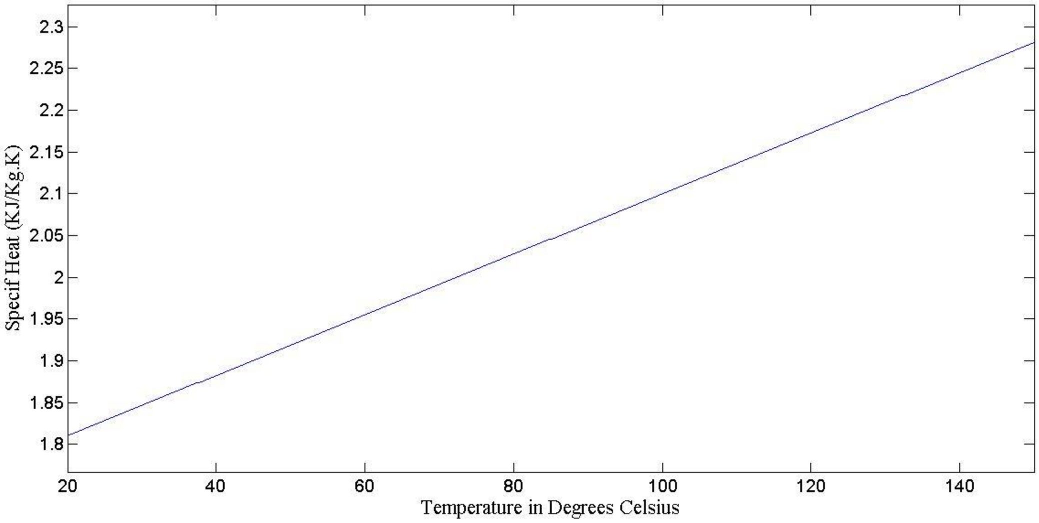


Figure A.2 Lubricant specific heat in (KJ/Kg.K) versus temperature.

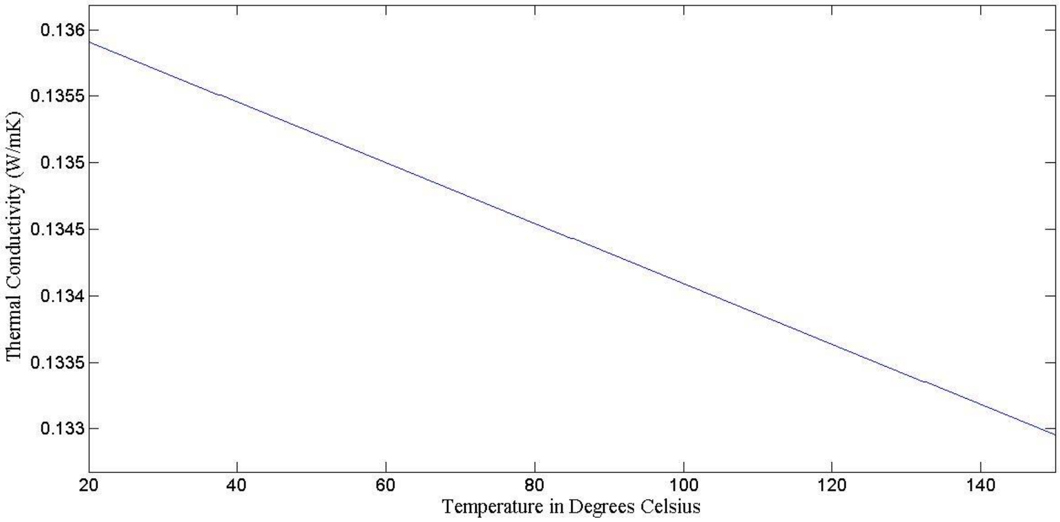


Figure A.3 Lubricant thermal conductivity (W/mK) versus operating temperature

A.2 Pressure Viscosity Relationship

The viscosity is the lubricant most important property that affects lubricant performance. The viscous drag friction torque and heat transfer in the bearing are also affected by the lubricant viscosity. The kinematic viscosity of the grease lubricant is that of the oil included in the mixture that is soaped to fabricate the grease. The viscosity is highly affected by two factors; the working pressure and temperature. One of the very famous formulas that are used to calculate the lubricant viscosity at different temperatures and different loads is the Roelands formula (Harris, et al., 2006).

$$\frac{\log_{10}\ \eta + 1.2}{\log_{10}\ \eta_0 + 1.2} = \left(\frac{T_0 + 135}{T + 135}\right)^{S_0} \left(1 + \frac{P}{2000}\right)^{z} \tag{A-4}$$

where η_0 is the kinematic viscosity in centi-stoke at the ambient temperature, T_0, in degree Kelvin, η and T are respectively the viscosity and temperature at running condition while P is the applied pressure in kgf/cm^2. The formula includes two exponents; S_0 and z can be found for each type of lubricant using the following two equations (Stachowiak, et al., 2014):

$$z = \frac{\alpha_{pv}}{5.1\times10^{-9}\left[\ln\eta_0 + 9.67\right]} \tag{A-5}$$

$$s_0 = \frac{\beta\left(T_0 - 138\right)}{\ln\eta_0 + 9.67} \tag{A-6}$$

In the above two equations α_{pv} is the pressure viscosity coefficient which is a lubricant property . that can be considered a property for each lubricant type and can be found in many tribology references such as (Stachowiak, et al., 2014; Stolarski, 1990).

Then Equation (A.4) becomes:

$$\frac{\log_{10}\ \eta + 1.2}{\log_{10}\ \eta_0 + 1.2} = \left(\frac{T_0 + 135}{T + 135}\right)^{s_0} \tag{A-7}$$

Using two known values of lubricant viscosity as indicated in grease and base oil data sheets (Lumax, 2011; US Enviromental Protection Agency, 2011) the exponent S_0 can be calculated as follows:

when considering T_0= 20 degree Celsius, v_0=40 cSt (centi-Stoke) and knowing that at T= 40 º C the viscosity of the lubricant v =19 cst, then substituting into Equation (A.7) will result in:

$$\frac{\log_{10}\ 19 + 1.2}{\log_{10}\ 40 + 1.2} = \left(\frac{(20+273)+135}{(40+273)+135}\right)^{s_0} \tag{A-8}$$

By solving the above equation, it is found that for the current lubricant $S_0 \approx 2.66$ and the formula to calculate lubricant viscosity at any temperature will be as follows:

$$\frac{\log \eta + 1.2}{\log 40 + 1.2} = \left(\frac{428}{T + 135} \right)^{2.66} \tag{A-9}$$

The above formula is used to draw the curve that expresses the lubricant viscosity at any temperature as shown in Figure A.1

A.3 Shear Rate Effect

Although it was firstly thought that the grease lubricant will behave just like the base oil, it is found later that the Non-Newtonian behavior of the thickener can change the lubricant rheology at different shear rates. For different lubricants the shear rate can cause two opposite phenomena, shear thickening and shear thinning. The change of the grease viscosity due to the increase of shear rate is of the first type, i.e. for higher shear rates the effective viscosity of the grease start to decrease till reaching its final value that equals the base oil viscosity. Two of the famous models for grease shear rate effect on viscosity modeling are; the Herschel-Bulkley model and Palacios model (Palacios, et al., 1984).